THE RELEVANCE OF THE SOCIAL SCIENCES TO ARTIFICIAL INTELLIGENCE AND EXPERT SYSTEMS

A study submitted in partial fulfilment
of the requirements for the degree of
Master of Arts in Librarianship
at the University of Sheffield

Mirko Popovic

September 1986

ABSTRACT

This dissertation analyses the need for a closer association between the social sciences and artificial intelligence (AI) and expert systems. The arguments are based upon a review of discussions from a variety of sources about AI and expert systems which reveal at present a wide variation in the descriptions of the current state of the art.

The discussion starts with the notion of interpretative flexibility of the term "intelligence" which consequently leads to disagreements about the research achievements of AI. This is followed by a description of the main issues in AI research and its development which also indicates that AI is a big, and increasingly growing business.

In the context of this increased interest in AI research and the discrepancies in reports, the section on sub-areas of AI and on AI as an interdisciplinary field reveals a narrow view of the role of the social sciences. On the one hand, it is often thought that the development of cognitive psychology is the only relationship between the social sciences and AI, and, on the other hand that social sciences are only concerned with the effects of AI, but not with its genesis. In contrast to this restricted view, knowledge, language, intelligence, etc., are defined as social concepts and the need for the social sciences' involvement with these main AI issues is stressed.

This relationship is illustrated through the example of expert systems. Throughout a discussion about the main issues in expert systems, wide variations in the assessments of the field are presented, together with a indication of the oversimplified and atheoretical approach, mostly found in "popular" literature. This is also a starting-point for the analysis of the fundamental problems in expert systems building, relevant to the social sciences, i.e. knowledge acquisition, knowledge representation, and explanation facilities. The main conclusion is that the social sciences will only be able to assess the impact and effects of AI and expert systems in different environments if they are also involved in research into these fundamental problems.

The final chapter discusses the relevance of AI and expert systems research to library/information systems. New methods of organizing and representing information in databases, and expert intermediary systems are identified as two areas which could benefit from such research. This also has important implications for library/information education.

To Breda

CONTENTS

ACKNOWLEDGEMENTS

The author would like to thank Professor M.F. Lynch of the Department of Information Studies, University of Sheffield, for his guidance and help throughout the production of this dissertation.

Thanks are owed also to Mr. H.M. Collins, University of Bath; Mr. A.S. Pollitt, Huddersfield Polytechnic; Mr. R. Davies, Exeter University Library; and Mrs. L.C. Smith, University of Illinois at Urbana-Champaign, who have supplied me with their latest papers and reports, relevant for my work.

Furthermore, I would like to thank Sheffield University Library for its inter-library loan service which provided me with many valuable articles.

Sincere thanks are offered also to my colleagues Amanda Saville and Carole Edwards for their time and effort in clarifying some expressions in my dissertation.

Finally, thanks are owed to the Culture Community of Slovenia and the National and University Library in Ljubljana for their financial support.

"... we need to investigate the relationship between the pronouncements of spokesmen on behalf of AI and the practical day-to-day activities of AI researchers."

(S. Woolgar, 1985, p. 567)

INTRODUCTION

> "... computerised searching services will not have their full impact upon user communities until direct user searching is widespread."
> (S. Pollitt, 1986, p.1)

There is an increasing tendency to bring computerized systems into the specialized domain of librarianship and information science: mechanized cataloguing and online information retrieval are two typical functions.

It was apparent as early as 1968, two years after the MARC-project was initiated, that,

> "A machine record is not simply a different physical means of recording a traditional bibliographic entry, for use in a traditional way" (Vickery, 1968, p. 1).

Has the library world considered the arguments for research aimed at the determination of the structure and content of an optimal bibliographic record in machine-readable form? I think not.

Computers have been used in libraries for approximately twenty years. Many of the larger libraries have automated circulation systems, online catalogues, serials control, financial and statistical reporting, etc. However, the new tools have mostly been used to perform the same type of work as before - witness the computerized catalogue card printer. Or, in other words,

> "... the present library systems have mostly been designed to the specifications of librarians, for librarians" (Hjerppe, 1983, p. 16).

The user, his access to the text and the knowledge in the text and the organization of knowledge using the new means available, have been neglected to a large extent. The only "library" organizations to experiment and utilize the inherent possibilities of machine-readable bibliographic representation have been the abstracting and indexing services.

But, however sophisticated these services might be, there are many problems, most notably in online bibliographic information retrieval systems. Although the rapid retrieval of references through online systems normally greatly reduces the amount of time spent searching for details of documents, it has little effect on other stages in the process of acquiring and using information. One of the greatest problems is that such access to knowledge does not indicate where the document carrying that information/knowledge can be found.

The second problem in bibliographic information retrieval is the specification of the user's requirements which relates to the procedures for classifying and indexing the document. However adequate index terms or thesaurus entries to a document may be, there is one main difficulty:

> "The problem is caused by the nature of the features: the fact that they are words. Words in isolation can have many meanings and when combined into phrases they can be cajoled into many subtle variations that cannot be easily simulated by logical connectives. Consequently, the user's specifications to an IR system can often return much irrelevant material, and attempts at refining the original request can result in no response" (Addis, 1982, p. 302).

Finally, untrained users can only search online databases having learnt artificial command languages, consulted manuals and relied on the help and intervention of trained information intermediaries. It is widely believed that end-users will be able to make their own requests online when search processes are simplified or made more "friendly".

These are some of the issues which have recently led to the increasing interest in library and information community in artificial intelligence (AI) and expert systems research.

The relevance of much of this research to library and information service area seems obvious. Walker (1981) put it in these words:

> "The information retrieval systems of information science and the knowledge-based expert systems of artificial intelligence can be viewed as constituting two ends of a continuum of facilities relevant for knowledge synthesis and interpretation. Considered in idealized form, both represent static states, the content of information retrieval systems providing the raw materials from which people derive information relevant for their needs; the expert systems embodying digested knowledge consensually validated as relevant for some area of inquiry" (p. 360).

One of the first attempts at bringing these two ends closer together can be found in Smith's article (1976):

> "In particular, information retrieval systems need no longer be limited in scope to the reference retrieval systems ..., but instead may be expanded to include fact retrieval, data retrieval, and question-answering as well" (p. 195).

An indication of the closer association between AI research and library/information science can be found in the development of intermediary expert systems, the main aim of which is to provide aid and assistance to users who wish to carry out their own online searches. Some of these systems are already available for public access, and others are being tested in experimental settings. On the other hand, although little progress has been made in developing expert systems for "traditional" library and information work, there are more and more articles which stress that the use of expert systems forces a rethinking of the methods of organizing and representing knowledge and information in order to make them more dynamic and interactive.

I do not intend to describe in detail the relevance of AI research for library/information science in this section - this will be done in the last chapter - but to stress the necessity of the library/information profession's awareness of these developments, and to put the whole area into a much broader context.

One of the main characteristics of AI research is the lack of a firm theoretical foundation which has its ground in the ill-defined term "intelligence", and consequently, in endless discussions whether machines can be intelligent or not. This, of course, leads to the discrepancies in reports about the achievements of the field, i.e., extraordinary optimism is countered elsewhere with claims that AI faces fundamental problems. Therefore, a complex understanding of the state of the art is needed when applying AI research results not only to library/information service area, but also to all other domains.

I would like to illustrate this statement by the example of the research project "Development of scientific and technical information in Slovenia 1986-90" (see Kornhauser, 1985) which emphasizes expert systems (also decision-making systems) as the final step of the development of a whole information infrastructure. This project, which can also be seen as an attempt to preserve artificial distinction between libraries and information services, proposes the following directions of development, as shown in Fig. 1.:

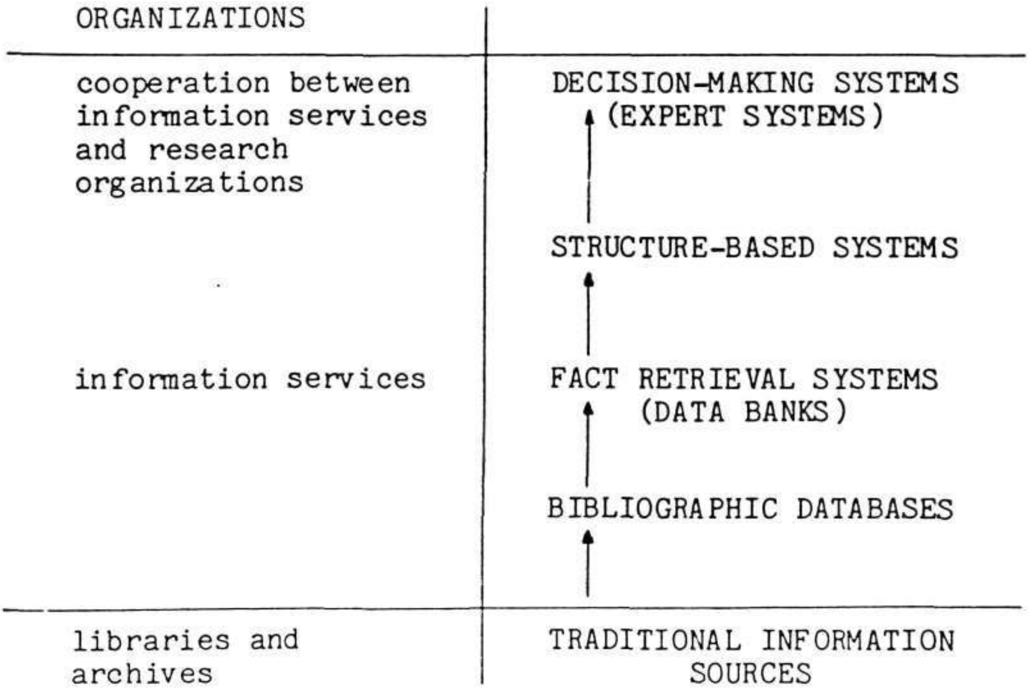
ORGANIZATIONS
cooperation between information services and research organizations
DECISION-MAKING SYSTEMS (EXPERT SYSTEMS)
STRUCTURE-BASED SYSTEMS
information services
FACT RETRIEVAL SYSTEMS (DATA BANKS)
BIBLIOGRAPHIC DATABASES
libraries and archives
TRADITIONAL INFORMATION SOURCES

Fig. 1. From information sources to expert systems (taken from the research project "Development of scientific and technical information in Slovenia 1986-90").

In addition, this research project does not take into account the different relationships between knowledge, communication, and information systems in the sciences, the social sciences, and the humanities. It is evident from Fig. 2. that this direction of development is <u>equally</u> proposed for the following complexes:

BIOMEDICINE

NATURAL SCIENCE
TECHNICAL SCIENCE
BIOTECHNICAL SCIENCE

ECONOMICS
SOCIAL SCIENCES
HUMANITIES

INTERDISCIPLINARY
COMPLEX

Fig. 2. Areas of application of research project <u>"Development of scientific and technical information in Slovenia 1986-90"</u>.

There is no doubt that the positive side of this project is in its introduction of "system-thinking" in information community, i.e., in linking bits of information into networks and showing the interrelationship between data, which is proposed in the development from automated bibliographic databases, factographic computer-supported databases, structured databases, to expert systems. Some problems which demand new approach in organizing information and knowledge in databases have already been described.

However, there are many questionable issues in this project, for example:

- uncritical adoption of "system-thinking" to all disciplines;

- artificial distinction between scientific and technical information services and libraries;

- unjustified reduction of libraries to the traditional role (libraries are even excluded from the automation, i.e., computerized bibliographic databases are only domain of information services).

Therefore, a much wider approach is needed, with a stress on the structure of knowledge as a central issue. This is important not only for library/information science, but also for the social sciences (especially sociology). In the context of this research project, there are many questions which should be of interest to both areas:

- what are the main differences in communication patterns, information seeking, structure of knowledge, etc., between sciences and social sciences, e.g., chemistry vs. sociology (1)?;

- how can knowledge in different disciplines be represented and formalized in expert systems?;

- are knowledge, language, intelligence, etc., as central issues of AI research, also social concepts, and thus not only the domain of hardware and software developers and so-called "knowledge engineers"?

This dissertation therefore aims to present a critical analysis of AI and expert systems research and a definition of the needs for the social sciences approach to AI and expert systems. The arguments below are based upon a review of discussions in a variety of sources about AI, expert systems, and their main issues. Two main characteristics of these discussions are evident:

1 - there is a wide variation in assessments and descriptions of the current state of the art;

2 - the relationship between the social sciences and AI research is reduced to cognitive psychology and linguistics; sociology is either excluded altogether or its contribution is only recognized in the discussions about the impact and effects of AI.

The AI perspective for library/information systems can only be clarified by the analysis of these features.

Therefore, to provide a framework for a closer association between the social sciences and AI research, and for a review of AI applications to the library/information service area, this dissertation begins with an analysis of the reasons for discrepancies in reports about AI. This is followed by the description of the main issues in AI research and its development, which will show that AI is a big, and increasingly growing business. This chapter will be concluded with the outline of the main sub-areas of AI and with a discussion on AI as an interdisciplinary field, where the reasons for the lack of a closer relationship between social sciences and AI will be examined.

To confirm these ideas, I will turn to examples taken from literature on expert systems which have been widely acclaimed as the applied end of AI research. Throughout the analysis of the main issues in expert systems and the description of their development, disagreements between the authors, and oversimplified and atheoretical approaches in the literature will be presented. On this basis, the fundamental problems in expert systems building, which are also relevant for the social sciences, i.e. knowledge acquisition, knowledge representation, and explanation facilities, will be analysed. Finally, library and information systems as potential domains for AI and expert systems will also be discussed.

Chapter 1

Artificial intelligence and the role of the social sciences .

> "Once intelligence has evolved to the level of knowledge based on language, its social aspect must surely dominate" (R. Stamper, 1985, p. 172).

1.1. Interpretative flexibility of the term "intelligence"

The idea that the digital computer will someday match or exceed the intellectual abilities of human beings has been put forward repeatedly since its invention. In the past thirty years a new discipline, called "artificial intelligence"(2) has emerged. It is said by Waltz (1982) that,

> "... computer programs written by investigators in artificial intelligence have demonstrated conclusively that in certain activities (including activities most people would say require intelligence, such as playing games) the computer can outperform a human being. ... At the same time the understanding of various features of human intelligence has been considerably enriched by the attempt to describe analogues of those features in the detail necessary for writing a program. As a result the analogy relating the performance of the computer to that of human intelligence has broadened and matured" (p. 101).

The goals of AI research are evident from these statements. One of them is development of computational models of intelligent behaviour. A more engineering-oriented goal is the development of computer programs that can solve problems normally thought to require human intelligence. These are very ambitious aims, and,

> "... neither has been achieved in any general sense" (Duda and Shortliffe, 1983, p. 261).

However, it can be said that few areas of research have been as exciting, promising, or bewildering as AI. After thirty years of use, the very name still has the power to provoke controversy.

In this context, it can also be said that AI is still a relatively young science, which is characterized by,

> "... the lack of any clearly defined way of carrying out research in the field" (Ritchie and Hanna, 1982, p.2).

The main problem in AI research, as shown by Ritchie and Hanna (1982) on the example of AM system (i.e., system which has been claimed to "discover" concepts and conjectures in elementary mathematics), is that published accounts often do not directly correspond to actual large and complex working programs. This means that,

> "... very little of AI research fits into the traditional "experimental paradigm" in which well-defined hypotheses are refuted by empirical investigations" (Ritchie and Hanna, 1982, p. 30).

Although the authors in the AI community agree upon the lack of firm theoretical and methodological foundations in AI research as one of major problems, serious research efforts in the last ten years have led to important achievements and to a substantial body of fundamental principles in AI(3). There is even consensus among researchers about the definition of AI, as provided by Barr and Feigenbaum (1981):

> "Artificial intelligence is the part of computer science concerned with designing intelligent computer systems, that is, systems that exhibit the characteristics we associate with intelligence in human behavior - understanding language, learning, reasoning, solving problems, and so on" (p. 3).

What is controversial, quite apart from the subject matter, is the name itself. Conflict in AI has been bound up with the focus on intelligence, and it is written by Fleck (1982) that,

> "Intelligence is not a socially or cognitively well-defined goal and every distinctive social group tends to have its own implicit definition, couched in terms of its own interest. Consequently research in AI has been oriented towards a variety of goals" (p. 172).

Indeed, there are no hard and fast criteria to decide whether a system is artificially intelligent or not. Some people hold the view that intelligence is an essentially human attribute, and that therefore "artificial intelligence" is a contradiction in terms. Others are convinced that, however clever computers become, they will never produce anything that is genuinely intelligent.

Many authors try to avoid such questions, for example Borko (1985), who said in one of his articles:

> "It is not my intent, nor is it necessary, to provide a very precise definition of artificial intelligence or to decide whether machines can think. We can leave these questions to the philosophers" (p. 105).

I believe, however, that understanding (not solving) these questions is very important from the following points of view:

- a lively debate centring on AI reveals some different positions on AI research (e.g. "AI is impossible" vs. "AI offers a way of humanizing technology", etc.) (4), and consequently, leads to wide variations about the achievements of AI research;

- one of the central points of the accepted definition of AI is that the design of intelligent systems is a multidisciplinary process, which can be summarized in the man - machine relationship. The role of the social sciences, which it might be expected to find the reasons for discrepancies in the field, is also found in this context. In other words, the social sciences should be aware of the interpretative flexibility of the term "intelligence", otherwise their function will be reduced to the analysis of the effects of AI in different environments. This would mean that the subject of research would remain in the hands of software and hardware specialists.

Therefore, the heart of the problem lies in the question "what counts as intelligence". On the following pages I would like to enlarge upon this question.

As it has been seen, the most widely accepted definition of AI is "designing intelligent computer systems, which exhibit the characteristics we associate with intelligence to human behavior". This merely imports the difficulty of using the word, in the human sense, to the technological sense. Michie and Johnston (1985) said that,

> "It is not altogether surprising that there is a problem with names here, since we have no sound definition of natural intelligence either. Some psychologists define it thus: "Intelligence is what intelligence tests measure". So what then are intelligence tests?" (p. 18).

It seems that the word "intelligence" is associated with endless ambiguities, or as emphasized by Aleksander (1984):

> "The construct is undoubtedly fuzzy, and one can justifiably question whether it is right to dub a technological area, as that of intelligent systems is intended to be, with this lack of precision" (p. 18).

However, it is interesting to stress that, at the same time, Aleksander (1984) tries to find a solution in the construct of intentionality which is for him a shift in the paradigm of intelligent systems, and is,

> "... very much a human construct and deals with our ability to relate to other people and objects, by understanding inwardly their likely behaviour. This is thought to be the key construct that will distinguish between illusory and real intelligent systems" (p. 10).

In this context, intentionality is also used in the meaning "knowing what one is talking about when referring to objects in the real world".

For some years now, a major philosophical debate has been taking place between leading scientists involved in AI on the question of whether current machines and programs, particularly those that process natural language, may be said to possess intentionality. According to Aleksander (1984), the major proponent of the notion that no currently built machine or program has intentionality is the American philosopher John Searle(5). Searle bases his argument on the example of the programs the aim of which is to simulate the human ability to understand stories. He introduces the problem with his famous Chinese Room:

> "Suppose that I'm locked in a room and given a large batch of Chinese writing. Suppose furthermore that I know no Chinese, either written or spoken..." (Searle, 1980, p. 418).

He further develops his idea that this English-speaking person, given perfect and copious memory facilities, could be given a mass of rules (in English) relating to the manipulation of Chinese symbols. Armed with this, he argues, this person is in the same position as a language-understanding computer, and would have as much success in answering questions submitted in sequence of Chinese symbols as the machine. However, no matter how successfully the job is performed, the performers of the task have no idea of the content of the story,

> "... the computer has nothing more than I have in the case where I understand nothing" (Searle, 1980, p. 418).

Searle makes his point by showing that the language handling routines in a language understanding program are clearly just rules for handling symbols.

Searle's major philosophical opponent is D. Dennett (1979), who believes that one could ascribe intentionality to many of the products of current work in AI. His argument starts with the simple notion that intentionality is a concept that falls into the class of beliefs and is not measurable in quantifiable terms. Therefore, if I believe that a machine possesses intentionality, this belief could be based on a need to have such a belief rather than on clear-cut physical evidence.

This argument is very similar to the one that the British mathematician, Alan Turing, used in the very early days after the invention and construction of the first automatic digital computer. In his paper "Computing machinery and intelligence", published in 1950, Turing addresses the question: "Can machines think?". He replaces an attempt to define meanings of the terms "machine" and "think" with a related question expressed in a relatively unambiguous words: "What will happen when a machine takes the part of A in this game?". The game referred to is the imitation game in which there are three participants: the machine (A), a human (B), and an interrogator (a machine and a person are in separate rooms, and both are interrogated by a third party via some sort of teletype set-up). The object of the game for the interrogator is to determine which of the other two is human and which is the machine. The machine "passes" the test if the interrogator is unable to determine the difference between the machine and the person.

This experiment is commonly called "Turing's test", which,

> "... effectively says that if a human is prepared to attribute human qualities to an object, then he endows it with those qualities" (Aleksander, 1984, p. 23).

I think that it is very important to stress that the concept of the attribution of human qualities to machines, as argued by Dennett and Turing, can lead to some serious over-simplifications. An illustration of this may be found in a program written by Ed Weizenbaum of the Massachusetts Institute of Technology (MIT) in 1966, called ELIZA. This program, which was largely designed to disprove Turing's notion of attribution, simulates a psychiatrist, encouraging the human "patient" to type in his troubles, in ordinary sentences, while the machine makes supposedly helpful comments in reply. A dialogue could look like this (6) - the computer's responses are printed entirely in upper case:

Men are all alike.
IN WHAT WAY

They're always bugging us about something or other.
CAN YOU THINK OF A SPECIAL EXAMPLE

Well, my boyfriend made me come here.
YOUR BOYFRIEND MADE YOU COME HERE

He says I'm depressed much of the time.
I AM SORRY TO HEAR YOU ARE DEPRESSED

It's true. I am unhappy.
DO YOU THINK COMING HERE WILL HELP YOU NOT TO BE UNHAPPY

I need some help, that much seems certain.
WHAT WOULD IT MEAN TO YOU IF YOU GOT SOME HELP

Perhaps I could learn to get along with my mother.
TELL ME ABOUT YOUR FAMILY

```
My mother takes care of me.
WHO ELSE IN YOUR FAMILY TAKES CARE OF YOU

My father.
YOUR FATHER

You are like my father in some ways.
WHAT RESEMBLANCE DO YOU SEE

You are not very aggressive but I think you don't
  want me to notice that.
WHAT MAKE YOU THINK I AM NOT VERY AGGRESSIVE

You don't argue with me.
WHY DO YOU THINK I DON'T ARGUE WITH YOU

You are afraid of me.
DOES IT PLEASE YOU TO BELIEVE I AM AFRAID OF YOU

My father is afraid of everybody
WHAT ELSE COMES TO MIND WHEN YOU THINK OF YOUR
  FATHER

Bullies.
DOES THAT HAVE ANYTHING TO DO WITH THE FACT THAT
  YOUR BOYFRIEND MADE YOU COME HERE
```

and so on.

The program is actually very simple. It looks for certain key phrases in the "patient's" statements, and for each key it has a present reply. In the above example the key phrases are "depressed", "my mother", etc. Sometimes the part of the patient's sentence following the key phrase is included in the machine's reply, with the pronouns and verb forms swapped round: "I" for "you", "are" for "am" and so on. Several other tricks - like associating keywords with a class or situation ("mother" implies "family") - help enhance the illusion of intelligent dialogue.

Although ELIZA's dialogue with the user appears surprisingly realistic, the program does it without having the slightest understanding of the content of what it is repeating. If you say to ELIZA, "Let's discuss paths toward nuclear disarmament", you might well get the nonsensical reply, "WHY ARE YOU TELLING ME THAT YOUR MOTHER MAKES PATHS TOWARD NUCLEAR DISARMAMENT", if you had introduced the word "mother" in your previous interchange. It is said by Michie and Johnston (1985) that,

> "ELIZA is nothing but a very carefully worked-out parlour trick. Weizenbaum intended it as a joke - a parody - and was appalled when established psychiatrists took it seriously and started talking about the possibility of automated psychotherapy" (p.25).

And indeed, many who encountered ELIZA attributed human properties of understanding, and even interest, to the simple computer program. Weizenbaum was shocked at the misinterpretation of his work and noticed three distinct results:

1 - a number of practicing psychiatrists seriously believed that such computer programs could grow into a nearly completely automatic form of psychotherapy (7);

2 - some people, conversing with ELIZA became emotionally involved with the computer;

3 - there was also a spread of belief that ELIZA demonstrated a general solution to the problem of computer understanding of natural language.

These conclusions have led Weizenbaum to discuss the dangers of work in the field, and his condemnation of excessive faith in technology can be found in his book "Computer power and human reason", first published in 1974.

I think that the above examples have showed that the quest for intelligent machines cannot rest only with the attribution argument. But why have I discussed the problem of intelligent machines without giving any final solution?

There is no doubt that these polemics among AI researchers reveal the interpretative flexibility associated with the notion of intelligence and intentionality. This feature begins to account for the variations in reports of the state of AI research, which can be found particularly in the example of expert systems, where,

> "... we might expect optimistic representations of the vitality, achievements and potential of the field from those involved in marketing expert systems" (Woolgar, 1985, p. 564-565).

And indeed, expert systems research is a field where the extraordinary optimism of some reports is elsewhere countered by considerable caution and pessimism about the achievements to date. On the one hand, expert systems are generally regarded as one of the most active areas of AI research, and on the other hand, there is considerable concern about the fact that the field currently faces fundamental problems (see, for example, Duda and Shortliffe, 1983; Leith, 1986).

What are the implications of the features of AI discourse outlined above? To recognize the interpretative flexibility of notions of "intelligence" is very important for the social sciences because,

> "... adherence to the view that the phenomenon for AI investigation are the inner processes responsible for "thought" and "intelligence" will place these entities beyond the reach of mere observational social sciences" (Woolgar, 1985, p.565)

Indeed, as it will be illustrated later, the role of the social sciences among AI researchers is too often seen only in investigations of the impact and effects of AI, or the input of social sciences in AI research is too often reduced on cognitive psychology. At this point, the main idea can be emphasized: the social sciences will be able to assess the impact of AI only if they also become involved in a detailed consideration of the processes of research activity in AI. Therefore, the social sciences should also be concerned with the genesis of AI, and not only with its effects. In this context I would like again to cite Aleksander (1984) who says,

> "... the crucial issue in assigning any form of human wisdom to a machine is that we must be able to understand plainly how this wisdom gets into the machine in the first place" (p. 25).

Wisdom - known in AI community under name knowledge is the central issue. To understand how human knowledge can be encapsulated (or encoded) into AI applications (e.g. expert systems) is the starting-point for a discussion about the impact and effects of AI. These discussions (e.g., AI and unemployment - see Partridge, 1986; how the status of the human expert will be affected by expert systems, legal implications of the use of expert systems, etc. - see Boden, 1984) cannot be realistic without this notion.

These ideas will be followed throughout my dissertation, with the main emphasis on knowledge representation, which will be analysed in the context of expert systems. But first, I would like to clarify the main issues in AI research and to give a short description of development of AI with the notion of AI as an increasingly growing business. On this basis, the problem of AI as an interdisciplinary field will be discussed, where the need for the closer association between the social sciences and AI will be stressed.

1.2. Some issues in the development of AI

AI had its origins in the late 1950s and early 1960s when it was recognized that electronic computers were more than giant calculators: they could process symbols, expressed as numbers, letters of the alphabet, or words in a language.

The second impulse for AI research can be found, according to Duda and Shortliffe (1983), in the shifting goal of much science which originally tried to obtain only quantitative descriptions of natural phenomena.

Unfortunately, not all natural phenomena can be expressed well in numbers. In particular, symbolic rather than numerical operations seem to characterize such activities as planning, problem-solving and deduction. Serious work on AI began when it was realized that computers as processors of symbols are potentially capable of being programmed to exhibit such intelligent behaviour. This non-numerical emphasis is a crucial characteristic of AI which distinguishes AI from the mainstream of computer science.

According to Newell (1983), there are two more factors that served to isolate AI within computer science:

1 - its choice of heuristic programming techniques, as distinct from algorithms favoured by computer scientists;

2 - its development of list-processing program languages, when the rest of computer science was moving toward the use of compilers.

What is a difference between heuristic programs and algorithmic programs? An algorithm is a precisely defined procedure consisting of a series of steps or program instructions for performing a specific task which would necessarily lead to a problem solution. In contrast, heuristic programs utilize approximate and exploratory methods based upon partial knowledge which might lead to the discovery of a problem solution but which could not be guaranteed to do so. Heuristics enable one to work with ill-defined problems. A classic example from AI research is the game of chess for which no algorithmic solution exists; heuristic rules could be programmed, for example, to choose between two moves by selecting the one that restricts the opponent's mobility to the larger degree. Heuristic programs incorporate procedures for selecting alternatives and evaluating the results of partial solutions while progressing toward a final goal.

List-processing languages, for example LISP (8), are computer languages that facilitate the processing of data organized in the form of lists. They transform a program statement into a sequence of machine actions. In contrast to these are compiler languages (e.g., COBOL, FORTRAN, PASCAL, etc.) which were originally used for numerical computations, and are transformed directly into machine language. Compiled programs can be executed with much greater speed, but list-processing languages allow a higher degree of interaction and user involvement, a matter which is one of the major concerns of AI researchers.

With regard to these issues, the initial work in AI limited itself to non-numeric but well-defined and well-constrained problems such as symbolic algebra, chess playing or game playing in general, puzzle solving, and simple theorem proving. According to Nilsson (1971), among the important techniques that emerged were general methods for representing information in symbolic data structures, general methods for manipulating these structures, and heuristics for searching through them.

But, as it is stressed by Sowizral (1985),

> "... most of the early work in artificial intelligence served only to tease researchers. It hinted at the feasibility of computer reasoning but fell far short of solving practical problems. Clearly, problem-independent solution methods could not handle the combinatorial complexity of real-world problems" (p. 180).

Therefore, AI researchers started asking how people solved real-world problems. A frequent answer was that people possess knowledge of which the programs were wholly innocent. Or, as it is said by Lenat (1984):

> "By the mid-1970's, after two decades of humblengly slow progress, workers in the new field of artificial intelligence had come to a fundamental conclusion about intelligent behaviour in general: it requires a tremendous amount of knowledge, which people often take for granted but which must be spoon-fed to a computer" (p. 152).

The central role of knowledge in intelligence explains why the most successful programs so far have been expert systems which operate in highly specialized domains. But to be efficient, this knowledge has to be combined with methods of conceptualizing and reasoning about the problem area. Or, in other words,

> "... in attacking a complex problem people draw on various methods - I call them sources of power - of using their knowledge of the world's regularities to constrain the search for a solution. They may invoke mathematical theorems or less formal rules of thumb; they may break up the problem into more tractable subproblems, or they may reason by analogy to problems that have already been solved. To the extent that computer programs already exhibit intelligence it is because they draw on some of these same sources of power. The future of artificial intelligence lies in finding ways to tap those sources that have only begun to be exploited" (Lenat, 1984, p. 152).

The growing recognition of the many kinds of knowledge required for high-performance reasoning systems changed the shape of AI research. Not surprisingly, research began shifting from the development of powerful and general but combinatorially expensive reasoning techniques to the development of effective techniques for representing large amounts of knowledge and effectively using that knowledge. In the words of Goldstein and Papert (1977),

> "Today there has been a shift in paradigm. The fundamental problem of understanding intelligence is not the identification of a few powerful techniques, but rather the question of how to represent large amounts of knowledge in a fashion that permits their effective use and interaction. ... The current point of view is that the problem solver (whether man or machine) must know explicitly how to use its knowledge - with general techniques supplemented by domain-specific pragmatic know-how. Thus, we see AI as having shifted from a power-based strategy for achieving intelligence to a knowledge-based approach" (p. 84).

The result of this shift was a rapid growth of AI research and its applications in a number of fields. The most successful programs so far, as I have already said, have been expert systems, development of which can be seen in the following rhetorical assertions: "The science of artificial intelligence ... is at last emerging from academic obscurity" (Evanczuk and Manuel, 1983, p. 139); "Expert systems provide 'practical uses for a useless science'" (Alexander, 1982, p.1); "Knowledge-based expert systems come of age" (Duda and Gaschnik, 1981, p. 238), etc.

Although expert systems will be analysed in a separate chapter, it is necessary at this point, because of their commercial implications, to illustrate the shift in AI paradigm by the example of the increasing interest in AI research, as expressed in the Japanese Fifth Generation Computer Systems Project, Alvey Programme in the U.K., etc.

1.3. From "General Problem Solver" to the Japanese Fifth Generation Computer Systems Project, or: AI as a big business

One of the characteristics of today's AI research is that the central core of research tools is applied in a bewildering and increasing variety of application areas - language understanding, expert systems, vision and robotics, to mention but a few. The multiplication of research areas - which will be described in one of the coming sections - accompanies and reflects a move from the great optimism in the early days of AI research during the 1950s and early 1960s, to a period of stagnation in the early 1970s, and to an explosive growth of interest in AI in the early 1980s.

Early optimism was evident, for example, in the attempts to build a "General Problem Solver" (see Ernst and Newell, 1969) which could deal with any area of knowledge. After many years of effort, reflecting the extent and complexity of the areas of knowledge concerned, this approach proved fruitless and the goal of a general intelligent inference mechanism was abandoned.

In the early 1970's AI research had been forced into the background. The situation was especially interesting in the U.K., where, according to Manchester (1986), the Lighthill report (under the auspices of the Science Research Council) halted AI research by recommending that government funding should be curtailed.

The report said that the lack of a bridging technology between theoretical, computer-based research in automation and research into neurobiology and psychology was a barrier to progress. Lighthill wanted results and to bring research to the point where it could generate something resembling a commercial product. Otherwise, the report noted, there were,

> "... doubts about whether the whole concept of artificial intelligence as an integrated field for research is a valid one" (cf. Manchester, 1986, p. 38).

As a result of this, AI research in the 1970s was continued in a much lower key, with more realistic goals, but with deeper ideas and methods and with better programming tools. One of the signs of maturity was the already mentioned recognition that knowledge is as important as reasoning.

In the beginning of the 1980s, AI was suddenly again in the centre of attention, not only in the academic world and computer corporations but, significantly, national governments, especially in the USA, UK, and Western Europe. There are two main reasons for this tremendous turning-point:

1 - successful development of expert systems and their commercial potential;

2 - launching of the Japanese Fifth Generation Computer Systems Project which was based on the estimation that commercial success lies in the use of AI methods and tools for the fifth-generation computers.

The fact is that the Japanese Fifth Generation initiative, at an international conference in Tokyo in the Autumn 1981, helped to crystallize an interest which had already been growing in the West, but placed the development targets well beyond those that most Western researchers would have set for themselves. The report of the Japan Information Processing Development Center - the Jipdec Report - stated the target in these words:

> "The Fifth Generation Computer Systems will be knowledge information processing systems having problem-solving functions of a very high level. In these systems, intelligence will be greatly improved to approach that of a human being" (cf. Bramer, 1985, p.3).

In their well-known book, Feigenbaum and McCorduck (1984) added that,

> "... the Japanese expect these machines to change their lives - and everyone else's. ... Their Fifth Generation plans say unequivocally that the Japanese are the first nation to ... have acted on a truth that has been emerging and reiterated for nearly two decades. The world is entering a new period. The wealth of nations, which depended upon land, labor, and capital during its agricultural and industrial phases ... will come in the future to depend upon information, knowledge, and intelligence"(p. 14).

But the concept of the fifth-generation computer systems, as illustrated by the widely circulated figure on the next page, i.e. Fig. 3., presupposes an implementation of AI-based subsystems with capabilities beyond those feasible in research today. However, this project is a good example of what has been considered to be within reach in terms of the application of AI-based concepts. There is no doubt that many of the themes for fifth generation computer systems arise from prior research in AI. While this diagram may seem confusing, it clearly shows the determination to move in the direction of intelligent knowledge-based systems. The Japanese are also talking about the main programming of their machines being carried out in a "logic programming language", a technique originating from AI work and differing radically from conventional computer languages. And this software should run on fifth generation hardware which will feature parallel processing in contrast to current sequential machine architectures. The interest of the Japanese government to support such research can be illustrated, according to Clarke and Cronin (1983), with the £200 million which were provided for this ten-year project.

It has already been said that the Japanese project sounded the alarm in the Western computer industry. One of responses was the British Government's Alvey Programme (9) for Advanced Information Technology. With the report of the Alvey Committee in 1982, AI was "rehabilitated" in official circles in the U.K., under the

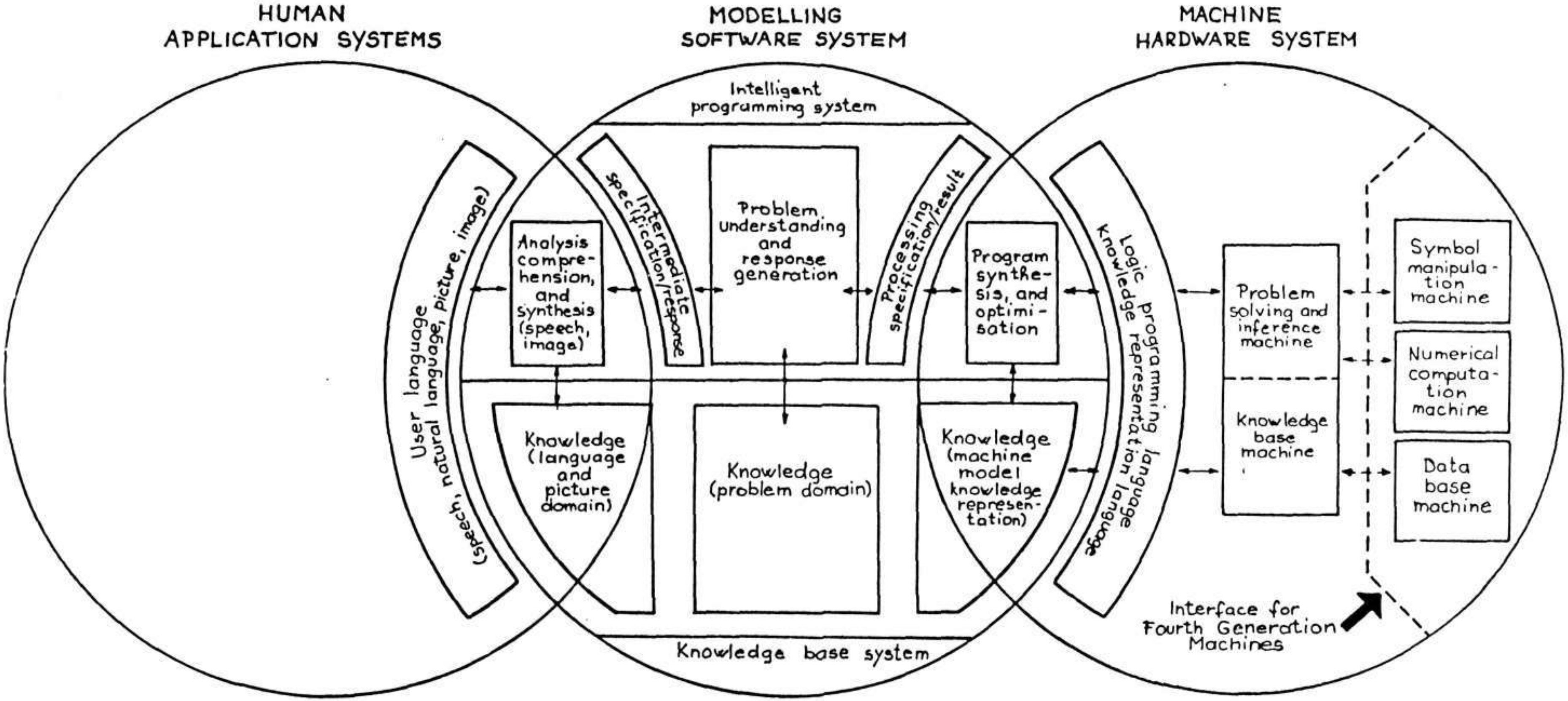
HUMAN
APPLICATION SYSTEMS
MODELLING
SOFTWARE SYSTEM
MACHINE
HARDWARE SYSTEM
Intelligent programming system
User language (speech, natural language, picture, image)
Analysis comprehension, and synthesis (speech, image)
Intermediate specification/response
Problem understanding and response generation
Processing specification/result
Program synthesis, and optimisation
Logic programming language
Knowledge representation langage
Knowledge (language and picture domain)
Knowledge (problem domain)
Knowledge (machine model knowledge representation)
Knowledge base system
Problem solving and inference machine
Knowledge base machine
Symbol manipulation machine
Numerical computation machine
Data base machine
Interface for Fourth Generation Machines

Figure 3. Conceptual diagram of the Fifth Generation Computer Systems (after JIPDEC)

new name of IKBS (Intelligent Knowledge-Based Systems). It is amusing to note the comment in the Alvey Report, after the highly damaging Lighthill Report, that,

> "The need to train additional personnel ... is particularly pressing in the IKBS area, where there are at present few active participants" (cf. Bramer, 1985, p. 1).

The Alvey Programme (10) is defined as a joint venture between three UK Government Departments - the Department of Trade & Industry, the Ministry of Defence, and the Department of Education and Science - and British Industry and academia. It is a five-year programme, begun in 1983 and costing £350m, of which £200 million comes from public funds and £150 million from industry. Its objective is to stimulate British information technology research through a programme of the following projects:

- Intelligent Knowledge Based Systems (IKBS);
- The Man-Machine Interface (MMI);
- Software Engineering;
- Very Large Scale Integration (VLSI);
- Computing Architectures.

The main interest for us is, of course, IKBS which is divided into the following four sub-programmes:

- IKBS Demonstrators;
- Research Themes, Projects and Clubs;
- Support Infrastructure;
- IKBS Awareness.

The demonstrator projects are intended to apply the ideas and techniques developed by the research community in systems viewed by industry as the precursors of market products. There are four IKBS demonstrator projects, all featuring the use of IKBS techniques in some specific area of industry, i.e., mechanical engineering, chemical industry, systems engineering, and applications in the data processing industry.

All research work in IKBS is organized within the framework of research themes and clubs, which have a common policy on such matters as tools and standards for languages. The structure for the year 1985, together with its financial implications is shown in Table 1.

Name of club	No of projects	Cost (£ mil.)	Research themes
Knowledge Based Systems	24	6.4	IKB Demonstrators Large demonstrator projects (IKBS components) Expert systems Intelligent front ends Intelligent computer-aided instruction
Logic Based Environments	6	0.5	Declarative languages Inference and knowledge representation
Declarative Architectures	19	16.7	Parallel architectures Intelligent database systems
Speech and Natural Language (jointly with MMI)	8	0.8	Natural language
Vision (jointly with MMI)	4	1.8	Image interpretation

Table 1. Research themes, projects and clubs in the IKBS for the year 1985 (after Alvey Programme, Annual Report 1985)

The need to bring to the attention of a wide spectrum of UK organizations the potential future importance of IKBS techniques, in general, and expert systems in particular, also led to the setting up of an IKBS awareness program and to the formation of several industry specific expert systems clubs. Typically, each club has 20 industrial organizations as members and the funds accumulated in this way are used to commission the building of expert systems in the area of interest to the club. At this moment, these clubs cover the following areas: real-time process control, insurance, transport industry, econometric modelling, data processing, and computer system fault diagnosis.

Another response to the Japanese Fifth Generation Computer Systems Project came in 1982 from the Commission of the European Communities which stressed the need for a European Strategic Research Programme in Information Technology - ESPRIT - based on the argument that information technologies represent the fastest growing sector of industry today.

Finally, as an indicator of the growing interest in AI research in the US this table is presented taken from Hayes-Roth (1985a) which estimates the levels and rates of change of some key technology measures:

Item	1984 level	1984-1985 change
Knowledge system prototypes under development	70	+50%
Knowledge systems being deployed	15	+100%
Knowledge systems being maintained	10	+200%
Knowledge engineering departments established	15	+150%
Senior knowledge engineers	40	+50%
Knowledge engineers	150	+50%
Knowledge engineer trainees	300	+100%
Applied AI graduate students	250	+20%
Undergraduate students in AI	2000	+50%
LISP or PROLOG programmers	2000	+50%
LISP or PROLOG installations	400	+100%

Table 2. Estimated measures of current US technology capacities (after Hayes-Roth, 1985a, p. 22)

The greatest benefit of the explosive growth of interest in and work on AI and expert systems can be seen from these examples, i.e., the much greater openness of commerce and industry to the ideas, techniques and tools of AI and the far greater willingness to experiment with building systems of their own. According to some of the latest studies (see Manchester, 1986, p. 38) the world market for AI products reached $342 million in 1985 compared with $181 million in 1984. These studies forecast that the market will be worth $665 million by next year. About half of that figure will be spent on software and 35% of the software market will be for expert systems products. On the basis of these figures it can be concluded that AI is a big, and increasingly growing business, at present most notably expressed in the fields of expert systems, natural language, and robotics.

As a result of this growing interest in AI it is necessary in the context of this study to give a short description of main sub-areas of AI research.

1.4. Sub-areas of the AI

When trying to outline the main AI sub-fields, there is again a problem of different views, and consequently a lack of firm theoretical foundations for AI. On the one hand, Aleksander (1984) identifies four major areas of AI, i.e., game playing, problem solving, artificial vision, and natural language understanding. On the other hand, in Fleck's article (1984), thirteen sub-areas of AI can be found.

To follow the main aim of this section, i.e., to outline some characteristics of AI sub-fields, help can, perhaps, be found in the so-called "AI-pie" - but not as the ultimate valid approach - as presented by Cercone and McCalla (1984) and shown on the next page (see Fig. 4.).

As it can be seen from Fig. 4., major efforts into AI research have concentrated on natural language understanding, computer vision, learning, theorem proving and logic programming, search, problem solving and planning, expert systems, knowledge representation, and other categories such as intelligent computer-aided instruction and tutoring, game playing, speech, automatic programming, and AI tools. At this point it is interesting to note that at present greatest attention in the AI community is paid to expert systems and natural languages. According to Smith (1985), papers submitted for the 1985 International Joint Conference on AI included 111 on expert systems, 99 on natural language understanding, and, interestingly for later discussion, only 4 on social implications.

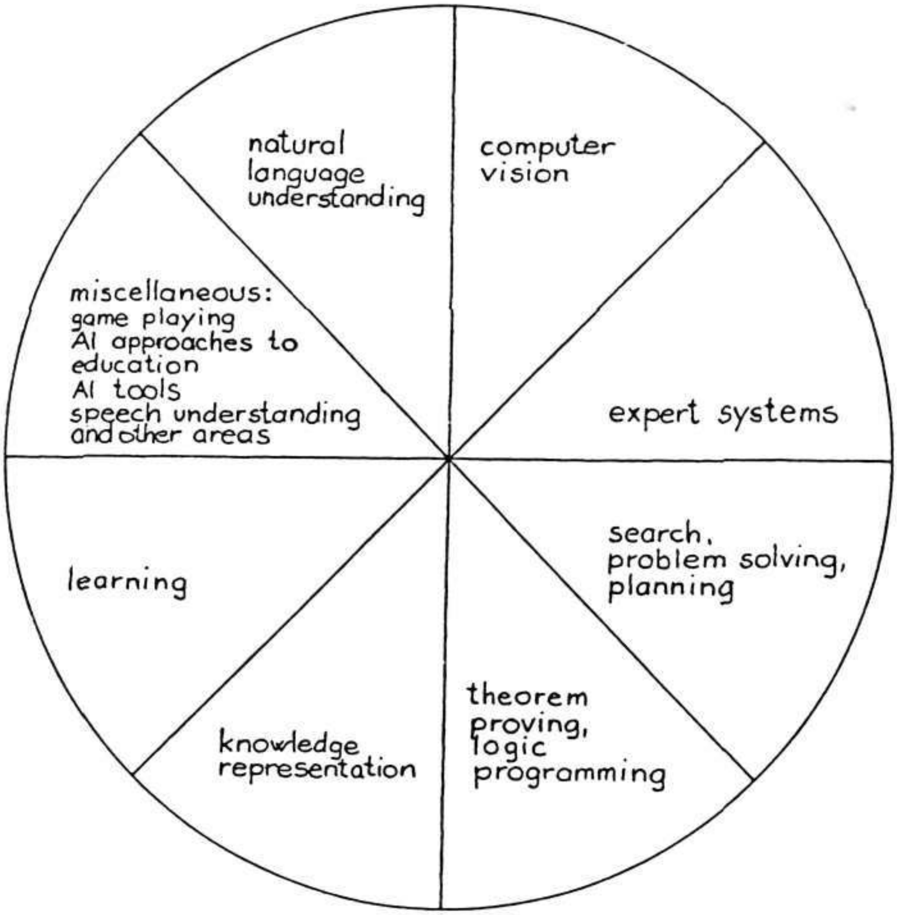
natural language understanding
computer vision
miscellaneous: game playing AI approaches to education AI tools speech understanding and other areas
expert systems
search, problem solving, planning
learning
theorem proving, logic programming
knowledge representation

Fig. 4. The "AI-pie" (after Cercone and McCalla, 1984, p. 281).

In the following sub-sections each sub-area in the "AI-pie" will briefly be outlined.

1. Natural language understanding: this has been one of the major research areas since the earliest days of AI. In the 1960s machine translation projects dominated, but they failed to account for meaning, context, etc. Winograd (1972) was the first to suggest a solution, by noting that conversations have to be about something. He suggested that the conversation should be about a restricted world. Further, he proposed that the extraction of meaning may be guided by a process of grammatical parsing.

According to Cercone and McCalla (1984), understanding natural language involves three levels of interpretation:

1 - syntactic processes "parse" sentences to make the grammatical relationships between words in sentences clear;

2 - semantics is concerned with assigning meaning to the various syntatic constituents;

3 - pragmatics attempts to relate individual sentences to one another and to the surrounding context.

The boundaries separating these levels are not distinct. At this moment the following directions of research into all levels of natural language exist:

- exploring alternative powerful parsing techniques;
- developing various schemes for explaining the semantics of natural language;
- modelling connected discourse and dialogue, especially focussing on pragmatic issues such as story structure, focus, reference, etc.;

- building natural language systems, e.g., front-ends to database systems.

One of the serious limitations of present-day natural-language systems is that they only work within a very limited domain of discourse. Their main advantage is that they enable a user to interact with databases without the use of specialized machine programs, for example so-called <u>question-answering systems</u> (e.g., LUNAR, developed in the early 1970's to allow lunar geologists to conveniently access, compare and evaluate the chemical analyses on lunar rocks and soil compositions accumulated by NASA during the APOLLO programme; and the LADDER system, which has been developed at SRI International and operates on a large naval database).

2. <u>Computer vision</u>: this is another very active, and very difficult area of AI research. Its basic objective is to interpret pictures (rather than to generate pictures which preoccupies computer graphics). Much research has been done into the problem of "pattern recognition", some of it with computers trying to make sense of television images of scenes consisting of simple geometrical objects: blocks, pyramids, boxes, etc. Sometimes the computer manipulates the object with a robot arm. Among the things the machine has to understand are that: the view of an object can be obscured by another in front of it; every thing must be supported by something or it will fall, etc.

Central to the problem is the fact that a picture contains an enormous amount of information. According to Michie and Johnston (1985), it is out of the question to do this processing with a conventional computer, because essential to the principle of such a machine is that it processes everything in a strict sequence, one item of data after another, and present circuits just cannot move fast enough. Owing to this "von Neumann bottleneck", for image processing in particular, information must unavoidably be processed many bits at one time, that is, in parallel. For machines to do this, a completely new type of hardware is needed.

Some machine vision systems for robots have, however, already reached the stage of being marketed (11). A very interesting situation emerges, if robotics is discussed from the commercial point of view. At the beginning of their development, there were great hopes of robots being general purpose machines. In the event, those hopes were soundly dashed. According to Fleck (1984), there were fewer than 200 units in use in the whole of the UK at the end of 1979, and, in general, diffusion everywhere was much slower than the manufacturers and promoters had expected, with only some 20-30 thousand robots in use worldwide by 1983. Practical experience has clearly demonstrated that certain robots are best suited to particular tasks within a narrow range of applications. At present, a differentiated set of more articulated and specific aims, with specialized knowledge and expertise developing around them, structures and guides research and development.

In fact, so far, industrial robots have been slow to diffuse, their economic feasibility has been difficult to demonstrate, and robot manufacturers have found it hard to achieve profitability. Despite this,

> "... excitement still prevails and there is much activity, with well over two hundred manufacturers in what is a relatively small market. ... This interest is based on the assumption that robots will be of great importance in the future. ... Robots have become a symbol of national technological progress, a sort of international virility symbol, to such an extent that many companies have already introduced them without concern for the economics, to prove to themselves and others that they can handle new technology" (Fleck, 1984, p. 208).

3. Search, problem solving, planning: it has already been said that the first big "successes" in AI were programs that could solve puzzles and play games like chess. Techniques such as looking ahead several moves and dividing difficult problems into easier sub-problems evolved into the fundamental AI techniques of search and problem reduction.

In general, there are three main problem-solving techniques:

- state-space search, which is nicely described by Nilsson (1982) through the example of the 15-puzzle. The main idea behind this kind of search is that we need to find a path from some initial state to any (one or more) goal state by applying operators to transform states into other states. This is also called forward direction of searching. State spaces can also be searched in a backward direction by starting with the goal state and applying the inverses of the operators to find a path to the initial state. Which approach is more appropriate depends on the particular problem and the nature of the state space;

- propagation of constraints: in this technique, the set of possible solutions becomes further and further constrained by rules or operators that produce "local constraints" on what small pieces of the solution must look like. More and more rule applications are made until no more rules are applicable and only one (or some other small number) possible solution is left. This process can be thought of as a type of state-space search that avoids the necessity of backtracking, since every existing solution must satisfy all the constraints produced by the rule applications;

- problem reduction: in this technique, the problem to be solved is partitioned or decomposed into sub-problems that can be solved separately, in such a way that combining the solutions to the sub-problems will yield a solution to the original problem. Each sub-problem can be further reduced, until "primitive" problems, which can be solved directly, are generated. Some decompositions of a problem may lead to solvable sub-problems, others may not. Problem solution is represented by a solution graph.

A typical example of an expert system which is based on problem reduction in conjuction with forward searching is DENDRAL (i.e., a computer system that proposes plausible chemical structures for molecules, given their mass spectrograms), which uses rules to narrow the searches to manageable numbers. This approach is called "reasoning by eliminating", and is based on early pruning, as illustrated in Fig. 5.:

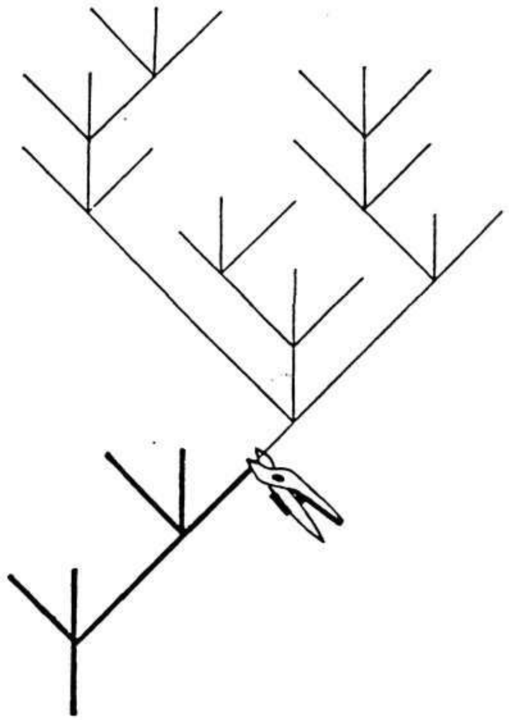

Fig. 5. "Pruning" a search tree. The shears indicate a place where the system could have grown a whole extra sub-tree in its search, but was saved the labour by the intervention of a pruning criterion which indicated lack of promise in that direction (after Michie and Johnston, 1985, p. 44).

An alternative approach, when the complexity of problem-solving methods increases, is producing plans.

Planning is preparing a program of actions to be carried out to achieve goals. An example is experimental planning in molecular genetics (see Stefic, 1981). A planner is required to construct a plan that achieves goals without consuming excessive resources or violating constraints.

The techniques described above underlie most areas of AI and are also used in expert systems building.

4. Expert systems: they are one of the most active and exciting areas of applied research in AI. Expert systems use AI problem solving and knowledge representation techniques to combine human expert knowledge about a specific problem area with human expert methods of conceptualizing and reasoning about that problem area.

Because of the central role of expert systems in my study they will be separately analysed in the following chapter. At this stage it can only be said that the work of the last few years has shown that programs which can operate at or near the level of human experts are feasible; several have been demonstrated as being capable of such performance in carefully selected, well-specified domains. As a result, the field is beginning to undergo a transition from basic research to applications, which has resulted in increasing commercial and industrial interest.

5. Theorem proving and logic programming: this area has also been significant in the development of AI

research. Theorem proving refers to the process of making logical deductions starting from a noncontradictory set of axioms specified in predicate calculus (firsts order logic). Robinson (1965) showed how it was possible to totally automate this process using a method called resolution. Theorem proving is also at the heart of the more recent development of the programming language PROLOG (12).

6. Knowledge representation: the central role of knowledge in building "intelligent machines" has already been stressed and there is no doubt that consequently,

> "... the representation of knowledge is the key issue in the development of AI" (Barr and Feigenbaum, 1981, p. 59).

But, surprisingly, although many representation methods have been developed in the last thirty years, the most important being logic, semantic nets, production systems, and frames, there is still no consensus on this topic. Many surveys of knowledge representation reveal a large variety of different views. Because of their importance, knowledge representation schemes will be outlined in the context of expert systems.

7. Learning: there are two very well known programs in "learning community", i.e., AM, and EURISCO. Lenat (1977) constructed a program (AM) that used heuristic search techniques to "discover" (although not prove) new concepts and theorems in mathematics from about hundred elementary

concepts in set theory. A follow-up program, EURISCO (Lenat, 1982), showed that similar methods could work in a wide variety of domains (e.g., fleet design, VLSI design, etc.).

This is a very important area of AI research because unless a computer can expand its own capabilities on the basis of "experience", the performance of the program is limited by the knowledge, foresight and available time of the programmer.

8. Other areas: there are a number of other areas that are often included in categorizations of AI research, including:

a) computerized game playing: interest in automating the game playing process has been manifested in AI since its inception as a field, not only for the obvious interest of getting a computer to play games well, but also because it was thought that the lessons learned by programming game playing strategies would generalize to the rest of AI. However, more recently, as AI programs have become more "knowledge intensive" , the so-called "weak, general" methods used in game playing have become less and less relevant. Although current game playing programs are extremely competent (e.g., there are game playing programs today that play near-master level chess), game playing as a research area is now pursued more for its intrinsic interest than for the lessons it can give to other areas of AI.

b) AI applications in education: there have been two main directions:

- producing intelligent tutoring systems that can behave with more subtlety and knowledge than traditional computer assisted instruction systems;

- developing learning environments for students, e.g., the LOGO programming system from which students can learn about programming and geometry.

c) AI tools: this field consists of developing programming languages (LISP, PROLOG), knowledge representation languages, and also of building specialized hardware (e.g., LISP machines). This area is of great commercial potential.

d) a number of other areas are also identified with AI, for example: speech understanding, automatic programming, etc.

I hope that this review of the main sub-areas of AI has provided some ideas about the comprehensiveness of the field. In short, research in AI can be characterized by the kind of activity or area of behaviour studied, or by the basic concepts and techniques that reflect the underlying mechanisms. In the first case, it is appropriate to refer to computer vision and robotics, language understanding, expert systems, etc. Considered in relation to the concepts and techniques, AI is concerned with issues of knowledge representation, search and problem solving procedures, logic programming, etc.

It can be concluded that this discussion has so far revealed some interesting features of AI research. One of them is that AI is not more the "property" of the academic world, but is becoming an important commercial field. As a result of the growing interest in AI, there are many calls in the AI community, addressed to the social sciences for investigations into the impact and effects of AI (see Davis, 1982; Boden, 1984).

At this point, I would like to emphasise the main argument once more: the social sciences will only be capable of assessing the impact of AI if they also have insight into key issues of AI research, for example, representation of knowledge. To testify this notion, it is necessary to open a discussion about AI as an interdisciplinary field and its relation to social sciences.

1.5. AI as an interdisciplinary field: the relationship between the social sciences and AI

Although there are again many difficulties in trying to define the multiplicity of AI roots, it seems that there is at least one consensus among AI researchers: AI is a part, although an isolated part, of computer science. The problem arises when trying to clarify the association between other disciplines and AI research. Minsky (1979), one of the pioneering researchers in this field, suggests that AI shares its goals with the following disciplines:

> "With computer science we try to understand ways in which information-using processes act and interact. With philosophy we share problems about mind, thought, reason, and feeling. With linguistics we are concerned with relations among objects, symbols, words, and meanings. And with psychology we have to deal not only with perception, memory, and such matters but also with theories of ego structure and personality coherence" (p. 400).

It can be said that this is a "classic" view of the interdisciplinarity of AI, shared among the majority of AI researchers which stresses the central role of computer science, philosophy, linguistics, and psychology.

In the context of the notion that knowledge representation is a central issue of AI research, the main interest lies in an analysis of the relationship between AI and the social sciences.

For many years it has been thought that the development of cognitive psychology was the only relationship between the social sciences and AI. The main idea behind this association was based on the argument about the validity of so-called "strong programme of AI". This programme relies on the adequacy of the "computational metaphor": a belief that the human mind can be studied as though it were a computer. For example, it is presumed that understanding speech involves computational processes in the brain that are similar to the processes performed by an AI program designed to accept natural language. It has been said by Gilbert and

Heath (1985) that,

> "The computational metaphor immediately suggests that AI and cognitive psychology have much in common, and indeed this has been for many years the perspective of the majority of cognitive scientists and AI practitioners" (p. 2).

There is nc doubt that AI research had a significant influence on the birth of cognitive psychology and thereby cognitive science during the 1970s. The AI concern with "intelligence" makes its close links with psychology unsurprising. For example, in her well-known book, Boden (1977) argues that AI,

> "... offers an illuminating theoretical metaphor for the mind that allows psychological questions to be posed with greater clarity than before" (p. 473).

However, as AI researchers have begun to tackle the difficult problems in understanding natural language, in representing knowledge and belief, in planning actions, and other areas, they have looked around at other disciplines to see how they have approached these issues. For instance, AI has been influenced by, and has in turn itself influenced, linguistics.

The association, therefore, between AI, on the one hand, and disciplines like psychology and linguistics on the other hand has been widely recognized and debated in terms of the implications of one for the other. But, sociology,

> "... whose interests clearly encompass language use and interaction, belief and knowledge systems, action and intentionality, is as yet unexplored territory" (Gilbert and Heath, 1985, p. 1).

Moreover, this restricted view of the role of the social sciences can also be seen in the use of the term "social" among AI researchers. For example, in Boden's article (1984) about AI and social forecasting the following statements can be found:

> "Put to commercial use, AI-programs will appreciably affect not only markets, but also personal and communal life-styles. Expert systems, for instance, will raise legal, social, and psychological problems of an unfamiliar kind. Whether they are used to make decisions or merely to provide expert advice to (probably less expert) decision-makers, the status of the human expert will inevitably be affected. And, on the international level, their use in countries lacking the relevant expertise may be seen ambiguously as helpful or exploitative - much as human technicians are" (p. 347).

In such views, "social" has to do with the effects of AI, but not with its genesis.

According to Woolgar (1985), some sociologists have also adopted a similar restriction of "social" in their treatment of AI. They are mainly interested in topics such as social attitudes to AI, public perceptions and acceptability of machine intelligence, and the likely effects of the implementation of AI in different institutional environments. At this point, the following question can be raised: can a sociologist without a detailed consideration of the process of the AI research itself (e.g., research into structure of knowledge and knowledge representation schemes) discusses the impact and effects of AI? I think not.

It can also be argued that this reduction of sociological capability has no legitimate theoretical background because it corresponds to the pre-Kuhnian view of science. Concomitant with that outdated view is a distinction between the "technical" (sometimes "intellectual" or "cognitive") aspects of science, on the one hand, and peripheral "social factors" on the other. This distinction was regarded as definitive of the scientific enterprise; "social factors" were precisely those factors not related to "science itself"; the domain of the "social" was regarded as outside or (at best) peripheral to the actual science. But recent work (i.e., post-Kuhnian sociology) has established that our understanding of science need not be so restricted; the nature and content of scientific knowledge is now recognized as a legitimate sociological object. Or, to put it into the AI context:

> "Sociological studies which focus solely on the impact of AI research, to the exclusion of the research activity itself, similarly underwrite the distinction between the scientific and the social" (Woolgar, 1985, p. 560).

With regard to the above arguments, two-sided exclusion of sociology from AI research can be discussed:

1 - AI researchers have not been interested in possible contributions of sociological research into crucial AI issues, for example knowledge representation;

2 - the function of the sociology has been reduced on the investigations of impact and effects of AI, but not the AI research activity itself.

With regard to the first notion, it can be said, however, that there are more and more articles in the literature, which argue strongly that the reduction of the relationship between AI and the social sciences especially the cognitive science (i.e., "computational metaphor") is misleading or mistaken. For instance, Coulter (1985) argues that it is confusing and inappropriate to use terms and expressions borrowed from computer science to explain human agency and social action.

The next opponent of the "computational metaphor" is Bateman (1985). His opposition to the cognitive science view, the goals of which are evident from Dennett's (1979) statement,

> "We want to be able to explain the intelligence of man, or beast, in terms of his design, and this in turn in terms of the natural selection of this design" (p. 12),

can be recognized in the following sentences:

> "The intelligence of man is, indeed, to be explained in terms of his design, but that design is not first and foremost the design of the biological entity. Human intelligence, perhaps as opposed to the intelligence of "beasts", is primary a social phenomenon - not one of the sub-personal psychology" (Bateman, 1985, p. 78).

Bateman suggests that some of the central topics of AI, such as knowledge representation and planning, that have so far been linked most closely with psychology, could more fruitfully draw on sociological investigations. Or, in his words,

> "In opposition to the view that knowledge representation has anything necessarily to do with sociology, I would like to suggest that knowledge representation ... is already and necessarily a sociological investigation. The main goal of ... cognitive science must be to articulate the "structures" and "processes" of the human life-world, not the processes of the hypothesised sub-personal psychological reality" (p. 65).

A similar approach can also be found in Stamper (1985) who argues that the way forward in AI research is to treat language and knowledge as predominantly social constructs:

> "Once intelligence has evolved to the level of knowledge based on language, its social aspects must surely dominate. ... However, in the world of AI, computational linguistics and cognitive psychology, language arises from some innate individual faculty for manipulating syntactic structures (p. 172). ... Adopting a simple, objectivist view of knowledge does eliminate many difficulties from knowledge engineering, but it does lead to people talking of knowledge as a kind of platonic substance which computers can process. ... Knowledge does not simply exist in a vacuum ; someone knows it" (p. 173)

In this context, it is very important to stress that Stamper rejects conventional logics, and proceeds to develop an alternative scheme, i.e., logic of action, which permits the handling of time, space, and context, issues which traditional logics have found difficult to deal with. Stamper (1985) emphasizes that,

> "Most mathematicians and logicians seem happily to concern themselves with a world of platonic reality where no one does anything, a world of timeless existence. More usefully perhaps, they deal with the rules for manipulating symbols and with the legitimacy of substituting one formula for another. But, when it comes to relating their paper and pencil formalisms to the world of practical affairs, mathematicians and logicians seem to do no better than their counterparts who program computers" (p. 174).

Stamper's notion lies at the "heart of the problem". The fact is that the techniques used for representing knowledge are relatively good at describing logical formulas and deductive necessities, and also hierarchies of objects. But coverage is particularly weak for the these other kinds of knowledge: time, space, events, and action. Therefore, the importance of developing new logics of this kind - Stamper's interest in this subject grew out of a practical study of how to design business information systems; i.e., in the world of practical affairs, where one judges the meaningfulness of information by its relationship to action - is that they may be used as the foundation for AI research that is more sensitive to the socially organized and public character of human action and cognition.

Consequently, this leads to a conclusion that all discussions about AI should derive from the analysis of the knowledge structure of the applied domains. To assess the potential of particular AI applications it is necessary to make a distinction between areas where knowledge can be represented in a highly structured way (e.g., chemistry, mathematics) and fields with "weak" formalism (e.g., different domains in social sciences). It is much easier to apply current AI techniques to the former domains, for the latter it is necessary to develop much more flexible methods. It follows that much more fundamental progress should be made in disciplines such as psychology and sociology of knowledge to understand the whole knowledge complex (transfer of knowledge, formalization of knowledge, common sense, etc.).

This claim, together with the argument that knowledge is a social concept, is already related to the second notion; or in other words, the distinctions between the "social" and the "scientific" which is a major barrier to a thoroughgoing sociological analysis of AI, "need to be transcended" (Woolgar, 1985, p. 559).

Therefore, it can be said that AI and sociology cannot profit from each other, as long as the relationship between AI and social sciences is reduced to a "computational metaphor", i.e., on the investigations of cognitive psychology into human learning and memory; and, as long as the contribution of sociology is seen to lie in discussions about the effects and impact of AI. This means that when sociology is asked to assess the effects of AI it should not only rely on the claims in AI literature, but should also investigate the practical day-to-day activities of AI researchers, for example, their approach to knowledge representation. I think that this notion also explains why there are so many discrepancies in reports about the state of the art.

In the following chapter, the need for a broader social sciences investigation into the main issues of AI (especially knowledge representation) will be explained by the example of expert systems where it is much easier to illustrate the concepts and techniques then in AI in general. Questions such as, can the knowledge of an expert be encapsulated in logical schemes, etc., will be raised, and terms such as "tacit" knowledge will be introduced. This will enable the discussion of, on the one hand, the potential and limitations of expert systems, and, on the other hand, the definition of some benefits deriving from knowledge representation research.

Chapter 2

Expert systems

"PROSPECTOR has discovered a molybdenum deposit whose ultimate value will probably exceed $100,000,000" (F. Hayes-Roth et al., 1983, p. 6).

"Unfortunately, this particular statement, which is similar to others we have encountered elsewhere, has no factual basis. ... PROSPECTOR's success to date has been scientific rather than economic" (R.O. Duda, P.E. Hart, R. Reboh, 1985, p. 359-360).

2.1. Variations in assessments of expert systems

It has already been said that the development of expert systems programs is one of the results of the shift in AI research to a knowledge-based approach. Expert systems are also known under the name knowledge-based systems. The fundamental assumption in expert systems is "knowledge is power" (Buchanan and Duda, 1983, p. 165), because the specific knowledge of the task (usually within the narrow area of expertise) is coupled with general problem-solving knowledge to provide expert-level analysis of difficult situations.

Unfortunately, it seems that the view on the development of expert systems, i.e., the shift from problem-independent solution methods to problem specific knowledge, is the only point of agreement between authors who try to describe and analyse this highly controversial topic. In all other aspects, starting from the basic issues, for example, definition, aims, architectural principles, practical use of expert systems, etc., it is very difficult to find any consensus. In addition, there is also considerable confusion in the terminology used to describe expert systems. This lack of consensus, of course, produces many difficulties when trying to define, describe, and assess this field. One of the most worrying questions is how the readers and potential users of expert systems can rely on these controversial statements. The greatest danger is in the uncritical adoption of claims from "popular" literature. To support the arguments about the wide variations in the field of expert systems, I would like to give some examples.

1 - disagreements about the definition of expert systems: a major problem when studying expert systems is the lack of clear definition of what they are. Bramer (1981) describes them as computer systems, which embody organized knowledge concerning some specific area of human expertise, sufficient to perform as skilful and cost-effective consultants. Sowizral (1985) sees them as computer programs modeled after human experts; they solve problems by mimicking human decision-making processes. Further, other authors (e.g., Denning, 1986) claim that expert systems are, after all, nothing more than computer programs. In contrast, many authors try to avoid definition of expert systems with explanations about what

distinguishes an expert system from an ordinary computer application (see Nau, 1983; Yaghmai and Maxin, 1984). I think that we could provide an endless number of such contradictory definitions.

Further problems are the preponderance of synonyms and the misunderstandings of basic terms. Expert systems are often referred to as knowledge-based systems, pattern-directed inference systems, problem-solving systems, etc. Two extreme examples are in the equating of expert systems, on the one hand, to rule-based systems (for criticism see Bramer, 1985), and, on the other hand, to intelligent knowledge-based systems (for criticism see Alvey Programme, 1985).

On this ground, one can agree with the statement, written by Cendrowska and Bramer (1984) who claim,

> "... no ... universally accepted theory exists at the present time, nor even a universally agreed definition of the term Expert System" (p. 229).

2 - <u>disagreements about evaluations of expert systems</u>: the fact is that expert systems have been widely acclaimed as the applied end of AI research, the long-awaited tangible outcome of research investment. Because the activity is finding a major <u>commercial market</u> (e.g., "One could imagine some use for expert systems in just about any sphere of business, engineering or research" - Webster and Miner, 1982, p. 60), it is often hailed as the justification and ultimate application of many years of endeavour in AI.

But the extraordinary optimism of some reports is elsewhere countered by considerable caution and pessimism about the achievements to date. There are some nice examples of these disagreements:

a) claims about DENDRAL: "The DENDRAL system is now in daily use by chemists at Stanford as well as by others in universities and industry" (Bramer, 1981), vs. "The DENDRAL and META-DENDRAL program are not used outside Stanford University and represent rather a demonstration of scientific capability" (Belkin and Vickery, 1985);

b) MYCIN, defined as most significant expert system, was never used by doctors for whom it was designed; ironically, this essential fact is very rarely mentioned in reports about expert systems;

c) R1, the system used to configure the VAX mainframe (DEC reports a $10 millions annual savings - see Hayes-Roth, 1985a), has only recently come under criticism for being much more difficult to amend than a straightforward program would be (see Leith, 1986);

d) PROSPECTOR is the subject of many claims about savings it had made for exploration companies (see Hayes-Roth et al., 1983); alarmed by these claims, the designers of PROSPECTOR wrote to Artificial Intelligence (see Duda, Hart, and Reboh, 1985) and pointed out, that, on the contrary, PROSPECTOR had made no savings at all. It had never even been used as the basis of any exploration plan.

These are only some of the examples which illustrate wide variations in claims about the field. These disagreements could be analysed on the additional topics, for example:

- what are the essential characteristics of expert systems (performance criteria or architectural principles)?;

- who are the users of expert systems (only experts or also naive end-users)?;

- what are functions of expert systems (interpretation, diagnosis, design, etc.)?;

- what is the aim of building expert systems (to duplicate intelligent human behaviour or only to assist experts)?, etc.

All these disagreements indicate that the field is still in a state which can be described as "pre-paradigmatic" in the terminology of Kuhn (1962), with many problems left to solve before expert system building,

> "... can emerge as a science rather than the craft it is now" (Bramer, 1985, p. 3).

At this stage the main task is to find the reasons which underpin these disagreements. But before undertaking such analysis, it is necessary to develop a complex view on expert systems, which will indicate an additional feature in expert systems descriptions, i.e., over-simplifications of the field.

The tendency for simplified descriptions derives mostly from research in AI, a field in which the complaint is often made that published accounts of research frequently do not directly correspond to actual working programs and often give a misleading impression of what has been achieved. This problem has already been mentioned in the context of the AM program (see Ritchie and Hanna, 1982).

Expert systems programs are often large and extremely complex, and thus not usually suitable for publication. Inevitably,

> "... published accounts tend to be in simplified form and this effect is compounded as second- and third-hand versions appear in textbooks and survey articles" (Cendrowska and Bramer, 1984, p. 230)

There is no need to stress the relevance of textbooks in passing on scientific knowledge to newcomers to the subject. To illustrate this notion the MYCIN program will be used, which will also serve as an introduction into the field of expert systems.

2.2. Oversimplified descriptions of main issues in expert systems (a MYCIN case study)

MYCIN is a rule-based expert system developed by E. Shortliffe (for detailed description of the system see Shortliffe, 1976; Cendrowska and Bramer, 1984) at Stanford University in the early 1970s. It was designed to assist physicians in the diagnosis and treatment of blood and meningitis infections. It tries to model the chain of reasoning used by the specialist by embodying his judgmental knowledge in the form of production rules, or in other words:

IF (condition)   THEN (implication)

It is claimed by Buchanan and Duda (1983) that, although MYCIN is now several years old, it is representative of the state of the art of expert systems in its external behaviour.

At this point, the first problematic question can be stressed: can a program which was never put into routine use in hospitals, be defined as representative of the state of the art?

MYCIN collects relevant information about the patient (e.g, his clinical conditions, symptoms, medical history or details of any laboratory findings) by conducting an interactive dialogue with the physician. It asks the basic questions first, for example (13) - user's responses follow double asteriks:

```
           ----------PATIENT-248----------
 1)  Patient's name:  (first-last)
 **  Dick H.
 2)  Sex:
 **  MALE
 3)  Age:
 **  52
 4)  Have  you  been  able to obtain positive
     microbiological information about a possible
     infection of Dick H.?
 **  YES
           ----------INFECTION-1----------
 5) What is the infection?
 ** ENDARTERITIS
```

It is claimed by Davis (1984) that this style of interaction is similar to what goes on in hospitals. But however good MYCIN's style of interaction and its explanation facilities (it can answer to how and why questions), when discussing the possible use of MYCIN in hospitals, it is said by Adler (1984),

> "I have not met a doctor yet, ... whose ego is small enough to handle that" (p. 262).

Can the answer to why it is so difficult to put medical expert systems into routine use be found here perhaps? Or it is necessary to analyse the structure of knowledge in MYCIN and see if all components of expert knowledge are encapsulated in the program?

The interview continues:

```
The first significant organism from this blood
culture (CULTURE-1) will be referred to as:

              ----------ORGANISM-1----------

9)  Enter the identity of ORGANISM-1:
**  Unknown
10) Is ORGANISM-1 a rod or coccus (etc.):
**  ROD
11) The gram stain of ORGANISM-1:
**  GRAMNEG
12) Have there been POSITIVE cultures yielding
    organisms about which you will NOT be seeking
    advice?
**  YES
```

Here, it is important to stress that MYCIN accepts "UNKNOWN" as a legitimate value. It does not need complete, or even correct, data for all answers. There are also many other features of MYCIN which cannot be illustrated here, for example:

- the system is tolerant of spelling or typing mistakes and can recognize synonyms;

- it communicates with the user in a subset of English.

As it will be seen, the information entered by the physician in response to these questions is used by the rules in an attempt to make a diagnosis. After between thirty and as many as eighty or ninety questions in one consultation, the physician sees a *diagnosis* similar to this:

```
INFECTION-1 is ENDARTERITIS with BACTEREMIA
<Item 1>  E. COLI [ORGANISM-1]
<Item 2>  SALMONELLA (species unknown)
                     [ORGANISM-1]
<Item 3>  KLEBSIELLA-PNEUMONIAE [ORGANISM-1]
<Item 4>  PSEUDOMONAS-AERUGINOSA [ORGANISM-1]
<Item 5>  ENTEROBACTER [ORGANISM-1]
<Item 6>  PROTEUS-NON-MIRABILIS [ORGANISM-1]
```

If, during this process, further information is required, the system will either try to infer it from the data it has already acquired, or it will ask the physician for it. As soon as a reasonable diagnosis can be made, MYCIN will compile a list of possible *therapies* and, on the basis of further interaction with the physician, will choose the most appropriate one for the patient. MYCIN prints out these comments:

```
[Rec 1] My preferred therapy recommendation is as
        follows:
      In order to cover for Items <1 3 4 5 6>:
         Give:  GENTAMICIN
         Dose:  128 mg (1.7 mg/kg) q8h IV (or IM) for
                10 days
         Comments:  Modify dose in renal failure
      In order to cover for Item <2>:
         Give:  CHLORAMPHENICOL
         Dose:  563 mg (7.5 mg/kg) q6h for 14 days
         Comments:  Monitor patient's white count

Do you wish to see the next choice therapy?
**  NO
```

In this case MYCIN recommended two medicines to treat all the possibilities.

When MYCIN is described and analysed in textbooks and articles it is often stressed that its most important features are as follows:

- a backward chaining inference system to reason "backwards" from diagnosis to symptoms;

- the use of rules with "certainty factors";

- an "explanation" facility to justify the inferences made by the system, combined with a means for the expert user to refine the system's knowledge base if deficiencies are found.

On the following pages I will illustrate some of the characteristics of MYCIN as they are presented in the "popular" literature, and at the same time, show how these descriptions are often over-simplified. In addition, to provide a complex view on expert systems, some other issues which cannot be encompassed in MYCIN (e.g., fuzzy sets), will also be outlined.

2.2.1. Architectural principles

The first of the over-simplified views on MYCIN can be found in the claim that its architecture is very simple, as illustrated in Fig. 6.:

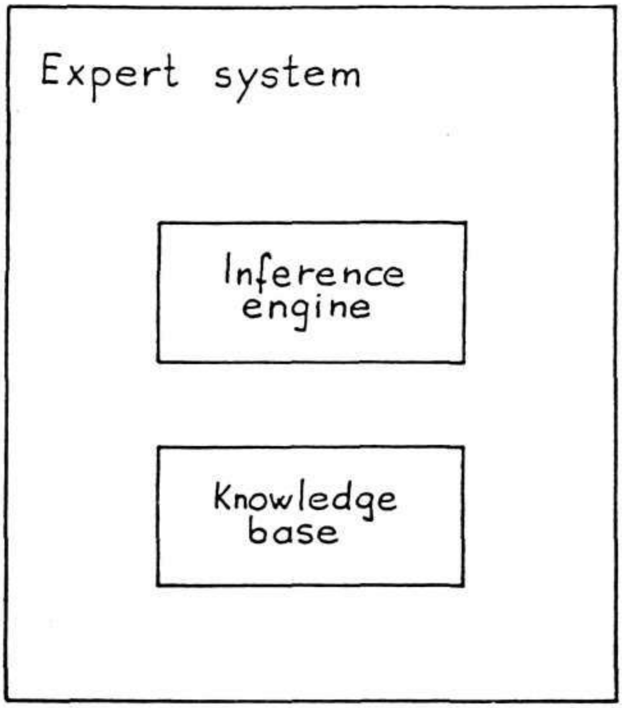

Fig. 6.: Structure of MYCIN (after Davis, 1984, p. 33)

According to such views, the knowledge base contains everything that is known about infectious disease diagnosis and therapy. The inference engine does the computation, taking knowledge from the knowledge base and putting it to work. Many authors also stress that in the case of expert systems in general it is important to think about the knowledge base and the inference engine *separately* because this should promote flexibility and transparency (i.e., the knowledge base can then be manipulated like any other data structure). In addition, the same inference engine can be kept even when a new domain requires a new knowledge. With this goal, as an extension of the MYCIN project, a subject-independent version known as EMYCIN ("empty MYCIN", or "essential MYCIN") has been set up, by removing the detailed medical information from MYCIN, whilst leaving the overall "backward chaining" framework, explanatory capabilities, etc., intact. With regard to these claims, it is necessary to sound a word of caution, and to outline some of the serious *over-simplifications*, starting with the *knowledge base*.

The knowledge base in expert systems is often defined as a body of knowledge specific to the problem area (e.g., meningitis) that the system is set up to solve; this knowledge is stored in some manipulatable form, by use of suitable formalism, or knowledge representation.

MYCIN's knowledge base, which contains about 450 rules, is often described like this:

Rule 085

If 1) the stain of the organism is gramneg, and
2) the morphology of the organism is rod, and
3) the patient is compromised host,

Then There is suggestive evidence (.6) that the
identity of the organism is pseudomonas.

The intention of such examples is to show that knowledge is contained in simple if-then inferential rules (also known as production rules), i.e., if we know that certain conditions have been met, than we can make certain conclusions. But, it is often "forgotten" to add that,

a) each rule has both an internal (stored) form and an external English translation. In the internal form, both the premise and the action part of a rule are held as a (LISP) list structure. The internal form of the above example is then as follows:

```
Rule 085

Premise ($AND (SAME CNTXT STAIN GRAMNEG)
              (SAME CNTXT MORPH ROD)
              (SAME CNTXT COMPROM T)
Action (CONCLUDE CNTXT IDENT PSEUDOMONAS TALLY .6)
```

b) although MYCIN's knowledge is largely rule-based, there is, according to Cendrowska and Bramer (1984) an important component, namely the creation of a list of potential therapies and the choice of the apparent first choice drug, which is algorithmic in nature.

c) MYCIN's architecture is much more complicated, and the entire MYCIN system comprises three subprograms (see Fig. 7.): the consultation program, the explanation program, and the rule acquisition program. It stores its information in two databases: a static database which contains all the rules used during a consultation, and a dynamic database which is created afresh for each consultation and contains patient information and details of the questions asked in the consultation to date.

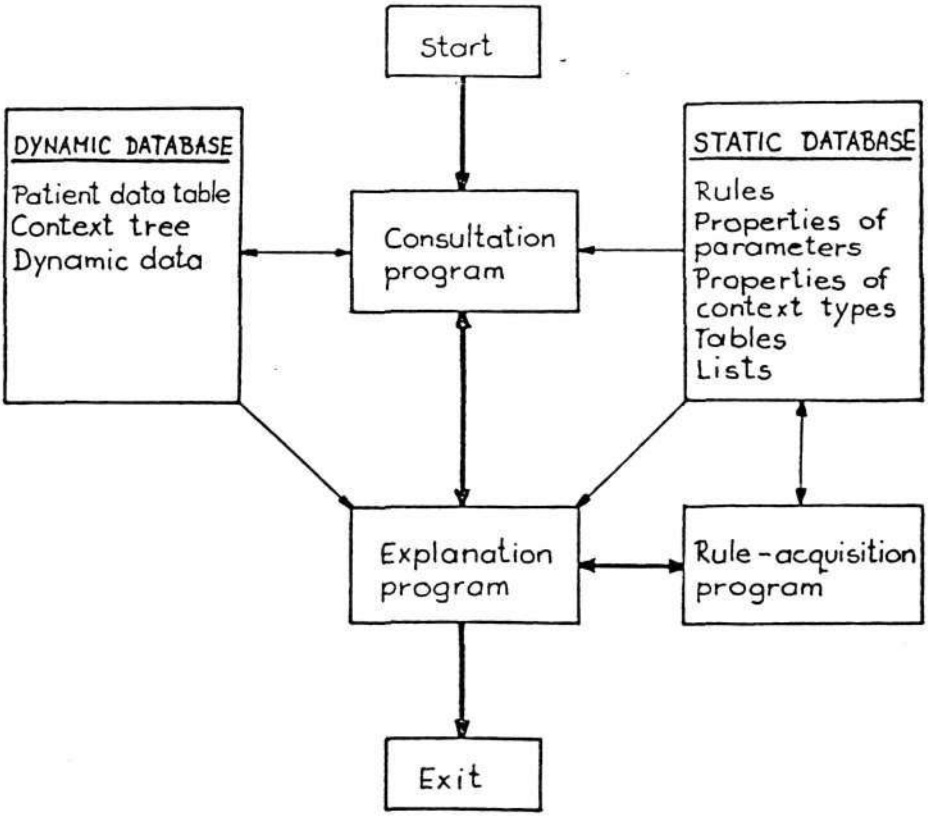

Fig. 7. Flow of control and information within the MYCIN system. Flow of control is indicated by heavy arrows, flow of information by light arrows (after Cendrowska and Bramer, 1984, p. 233).

2.2.2. Inference methods

The inference engine of expert systems makes decisions about how to use the system's knowledge by organizing and controlling the steps it takes to solve current problems. Thus, inference methods are closely coupled to knowledge representation schemes. According to

Buchanan and Duda (1983), data-driven control and goal-driven control are the two main control methods in rule-based systems.

With data-driven control rules are applied whenever their left-hand side conditions are satisfied. To use this strategy, one must begin by entering information about the current problem as facts in the database. Here it is assumed that a rule is applicable whenever there are facts in the database that satisfy the condition in its left-hand side. Data-driven control is also known by names "bottom up", "forward chaining", "pattern directed", etc. Its main advantage is in the quick response to input from the user. The potential inefficiency of this strategy is if that more than one rule is applicable, there is the problem of deciding which one to apply.

This problem can be avoided by using a goal-driven control strategy which focuses its efforts by only considering rules that are applicable to some particular goal. This strategy has also been used in many systems, and is variously known as "top-down", "backward-chaining", "consequent reasoning", etc. A primary virtue of this strategy is that it does not seek information and does not apply rules that are unrelated to its overall goal.

Data-driven and goal-driven strategies represent two extreme approaches to control. Various mixtures of these approaches have been investigated in an attempt to secure their advantages while minimizing their disadvantages (14).

The popular impression of MYCIN is that of a system with a simple backward-chaining control structure acting on a body of rules. For example, if the program was trying to deduce the identity of an organism (see the above example), one of the rules invoked would be Rule 085.

But the first problem with MYCIN control strategy is that interactions are too time-consuming, and this can be unacceptable when rapid, real-time response is required.

In addition, according to Cendrowska and Bramer (1984), there are many derivations from simple backward chaining, in particular the use of antecedent rules, self-referencing rules, mapping functions, etc.

2.2.3. Unreliable data or knowledge

One of the characteristics of an expert's work is "reasoning under uncertainty". Experts sometimes make judgments under pressure of time; all the data may not be available; some of the data may be suspect; some of the knowledge for interpreting the data may be unreliable. The general problem of drawing inferences from uncertain or incomplete data has led to a variety of technical approaches in expert systems building.

One of the earliest and simplest approaches to reasoning with uncertainty is the use of numbers, called certainty factors which indicate the strength of a heuristic rule. This approach was also used in the MYCIN system.

In the case of MYCIN a certainty factor is a number between -1 and +1 and is used to indicate the degree of belief that the value of the clinical parameter is the true value. A certainty factor of +1 indicates that the parameter is "known with certainty" to have that value. A certainty factor of -1 indicates that the parameter is known with certainty not to have that value. Certainty factors can be either computed or entered by the physician.

This is also one of the questionable features of MYCIN, although it is often claimed in the literature that handling uncertain information is one of the main advantages in expert systems building. According to Adams (1976), there are interdependence restrictions which need to be applied to the estimation of certain parameters ("measure of belief" and "measure of disbelief" in a hypothesis, supplied by the physician) to maintain internal consistency, but which are not included in the MYCIN model. In addition, the use of certainty factors as a means of ranking hypotheses is also suspect, since examples can be given of cases where, of two hypotheses, the one with the lower probability would have the higher certainty factor. On the basis of Adam's analysis, it would seem that the MYCIN model has serious limitations.

Another approach to inexact reasoning, very popular in the last few years, is based on the theory of fuzzy sets, first proposed by Zadeh (1979). In contrast to conventional set theory where the function "belongs to" assigns objects to sets, in fuzzy set theory a "degree of belonging" may be specified when making this assignment. For example, were one to define two sets "tall people" and "short people", given a person six feet tall one would assign him wholly (that is, using a multiplier of one) to the tall class. However, where he five feet nine inches tall the notion "quite tall" may be expressed by applying a belonging multiplier of 0.8 for the tall class and 0.2 for the short class. This multiplier notion should not be confused with the idea of a probability. There is no probability involved in saying that someone is quite tall. This approach is appropriate for areas where subtle distinctions are needed between objects.

The whole topic of reasoning under uncertainty is reviewed by Buchanan and Duda (1983).

2.2.4. EMYCIN as introduction to "shells"

It has already been said that EMYCIN was developed to provide a framework in which other systems can be built. The potential value of using EMYCIN is well illustrated by Feigenbaum (1979). In describing the development of PUFF, an expert system for the diagnosis of pulmonary function disorders constructed within the EMYCIN framework, Feigenbaum points out that the development time taken to reach a working system based on the analysis of some 250 test cases was less than 10 man-weeks of effort by knowledge engineers, with less than 50 hours of interaction with subject experts.

But, apart from the fact that using EMYCIN as a standard framework can greatly reduce the development time of an expert system, there are weaknesses which may often not be apparent to the user. One of the restrictions is the overall representation chosen (i.e., contexts, joined in a tree structure). Commenting on this representation, Van Melle (1980) remarks:

> "EMYCIN was not designed to be a general purpose representation language. It is thus wholly unsuited for some problems. ... The framework seems well suited for some deductive problems, notably some classes of fault diagnosis. ... It is less well suited for "formation" problems, ..., simulation tasks, and tasks involving planning with stepwise refinement" (cf. Bramer, 1985, p. 4).

Van Melle goes on to discuss the system's backward chaining form of reasoning and comments:

> "Backward chaining works well for simple deductive tasks, but as soon as a problem poses additional constraints on the means by which reasoning can or should be performed, more sophisticated control mechanisms may be required" (cf. Bramer, 1985, p.6).

There is no doubt that the choice of a suitable control mechanism is one of the key problems of AI programming. Therefore, the same comments should be made for all commercially-available "shells" which have appeared in the last few years and are based on either EMYCIN or its fairly close relative PROSPECTOR.

2.2.5. Explanation facilities

It is often claimed that explanation facilities are one of the greatest advantages of expert systems over ordinary computer programs (see Duda and Shortliffe, 1983). Michie and Johnson (1985) go even one step further and speak about the need to develop a human window as the opposite to the lunatic black box.

In the case of the MYCIN explanation facility, a physician can ask why the question the program asks is relevant. The response will be something like this:

```
14) Is the patient's illness with ORGANISM-1
    a hospital-acquired infection?
**  Why?
```

The user asks "WHY" to see an explanation of the line of reasoning MYCIN is following. The program first rephrases the question and then shows the rule whose premise it is trying to establish.

```
[i.e. WHY is it important to determine whether the
infection with ORGANISM-1 was acquired while the
patient was hospitalized?]

[2.0] This will aid in determining the identity of
        ORGANISM-1.
   It has already been established that
    [2.1] the morphology of ORGANISM-1 is rod
    [2.2] the gram stain of ORGANISM-1 is gramneg
    [2.3] the aerobicity of ORGANISM-1 is facul

   Therefore, if
    [2.4] the infection with ORGANISM-1 was acquired
           while the patient was hospitalized
   then
           there is weakly suggestive evidence (.2)
           that the identity of ORGANISM-1 is
           pseudomonas [RULE050]
```

MYCIN was using rule 50 and already had certain information about the patient. If the physician adds this fact, MYCIN will know whether this rule will fire. If it does fire, MYCIN will have more information about the topic that it is dealing with.

In addition to interrupting a consultation, the physician can wait until the end and ask MYCIN how it reached its conclusion:

```
** HOW DID YOU DECIDE THAT ORGANISM-1 WAS AN E.COLI/
I used rule 084 to conclude that the identity of
   ORGANISM-1 is e.coli. This gave a cumulative c.f. of
   (.51).
I used  rule 003 to conclude that the identity of
   ORGANISM-1 is e.coli. This gave a cumulative c.f. of
   (.43).
```

Therefore, by keeping track of the rules that were used, MYCIN can describe how it reached its conclusion.

As it can be seen, MYCIN is one of expert systems with quite dynamic explanation facilities. It is, perhaps, useful to stress that MYCIN's explanation capability was extended by the expert system GUIDON to include provision for tutoring.

However, there are more and more scepticisms about the adequacy of current explanation facilities, for example:

> "Explanation systems ... usually displayed the inference strategy of the system, not the expert. We did not explain our reasoning to other people in the way expert systems did. Explanation systems do not mirror how people talked to each other. ... Current explanation systems appeared to be good for finding out how the system had come to its decision, e.g., for debugging purposes. But as a method for explaining information to a naive user they were quite poor" (Alvey Programme, 1985, p. 117).

This discussion has so far shown, on the one hand that, when discussing expert systems, it is always appropriate to sound a note of scepticism, i.e., it is necessary to have in mind that this field is still ill-defined. Or, in other words, the foundation on which expert systems rest has been described by Sheil (1984) as being,

> "... a weak technology with few good boundaries and ... atheoretic" (cf. Town et al., 1985, Section 3.1., p.10).

On the other hand, it has also been seen that there are four main issues in the design of expert systems:

- knowledge representation;
- inference methods;
- methods for reasoning under uncertainty;
- explanation facilities.

Deriving from the idea of knowledge as the central issue in AI research, it will be useful to outline the main knowledge representation schemes.

2.3. Representation of knowledge in expert systems

Before saying anything about knowledge representation schemes I would like to emphasize that this section will be quite brief because it is very difficult not to become a "victim" of over-simplified approach discussed in the previous section.

It has been repeated many times in this study that the fundamental observation arising from work in AI has been that expertise in a task domain requires substantial knowledge about that domain. The effective representation of domain knowledge is therefore generally considered to be the keystone to the success of expert systems.

In expert systems building the following types of representation systems have been used:

- rule-based systems;
- frame-based systems (and semantic nets);
- logic programming systems.

Such frameworks are often called representation languages because, as with other programming languages, their conventions impose a rigid set of restrictions on how one can express and reason about facts in the world. At this point it is not my intention to discuss if the entire knowledge of an expert in a particular domain could be expressed in such languages, or to stress the importance of understanding the structure of such knowledge, but to provide a brief description of these three ways of representing knowledge(15).

2.3.1. Rule-based production systems

Rule-based production systems, developed by Newell and Simon (1972) for their models of human cognition, are defined as a modular representation scheme that is finding increasing popularity in expert systems. It is claimed by Hayes-Roth (1985b) that they constitute the best currently available means for codifying the problem-solving know-how of human experts. The basic idea of these systems is that the database consists of rules, called productions, in the form of condition-action pairs:

IF (condition)   THEN (action)

The left hand side of the rule describes a condition, and the right hand side describes the consequence if the condition is met. Once a database of such rules has been developed, it is possible to apply them systematically in a given context, in effect generating and testing hypothesis until one that applies is found.

As has been described, a typical example of a rule-based system is the MYCIN program which contains about 450 rules. Another two very well known expert systems which are based on production rules are R1 (configures the VAX/780 computer and contains over 2000 rules) and PROSPECTOR (aids geologists in evaluating mineral sites for potential deposits and contains about 1600 rules). In the case of R1, such a rule in a database may read (after Kraft, 1984):

If   The current subtask is assigning devices to unibus modules
     and there is an unassigned dual port disk drive
     and the type of controller it requires is known
     and there are two such controllers, neither of which has any devices assigned to it
     and the number of devices which these controllers can support is known

Then Assign the disk drive to each controller and note that each controller supports one device.

It is also claimed that production systems have two characteristics: first, existing knowledge can be refined, and new knowledge added, for incremental increases in system performance (derivations from this principle have also already been stressed). Second, systems are able to "explain" their reasoning.

However, the homogeneity and simplicity of expression attained in rule-based systems may, according to Fikes and Kehler (1985) reduce the ability to express other kinds of knowledge; in particular, their expressive power is inadequate for defining terms and for describing domain objects and static relationships among objects. These inadequacies can be handled by another knowledge representation technique, i.e, by frames.

2.3.2. Frame-based systems

Frame-based systems are the most recently developed AI knowledge-representation scheme. Frames are data structures in which all knowledge about a particular object or event are stored together.

Many different variations have been proposed for frame-based knowledge representation, but most of them include the idea of having different types of frames for different types of objects, with fields or slots in each frame to contain the information relevant to that type of frame. Thus a frame for a book description will have slots for title, author(s), publication date of the book, number of pages, etc. To describe a particular book, a copy of this book frame would be created, and the slots would be filled in with the information about the specific book being described.

One of the most often cited kinds of frames was developed by Schank (1975) in the context of a "theory of conceptual dependency". This, among other things, attempts to represent most events in terms of a small numbers of primitive actions. Each primitive action may be represented by a single kind of frame. For instance, Schank's theory casts "take" and "give" as two examples of the same phenomenon: a transfer of possession. The frame for a transfer of possession is:

```
name of frame:
type of frame:  transfer of possession
source:
destination:
agent:
object:
```

where the source is the person or thing from which the object is taken, the destination is the person or thing to which the object is given, and the agent is the one who performs the transfer. When the above frame is ready-made for the sentence "Bill took the book from Margaret", the result is:

```
name of frame:  T1
type of frame:  transfer of possession
source:  Mary
destination:  Bill
agent: Bill
object: book
```

The main advantage of frames is that all the relevant information is collected together, accessing and manipulating information is then easier. Several computer languages have been or are being developed to provide ways to manipulate frames, one of the most popular being KRL (see Bobrow and Winograd, 1977).

Frame representations are basically equivalent to semantic nets which were invented as an explicitly psychological model of human associative memory. Semantic nets are like frames in the sense that knowledge is organized around the object being described, but here the objects are represented by nodes in a graph and the relations among them are represented by labeled arcs. According to Mylopoulos and Levesque (1983) there are three advantages of semantic nets. Due to their nature, they directly address issues of information retrieval (this notion will be outlined in the last chapter). Another important feature is the availability of organizational principles. A third is the graphical notation that can be used for network knowledge bases and that enhances their comprehensibility. A major drawback of network schemes is the lack of formal semantics and standard terminology.

2.3.3. Logic programming systems

The key idea underlying logic programming is programming by description. In traditional software engineering, one builds a program by specifying the operations to be performed in solving a problem, that is, by saying how the problem is to be solved. In logic programming, one constructs a program by describing its application area, that is, by saying, what is true. A description of this sort becomes a program when it is combined with an application-independent inference procedure. Applying such a procedure to a description of an application area makes it possible for a machine to draw conclusions about the application area and to answer questions even though these answers are not explicitly recorded in the description.

One such language is PROLOG (see Kowalski, 1977; and Clocksin and Mellish, 1981) which works with objects and their relationships, specified by the programmer as rules. Relationships might be (16):

```
John likes Mary
Philip father-of Charles
Charles father-of William
Mary likes John
```

New relationships can be defined:

```
x friends-with y if x likes y and
                    y likes x
```

and questions can be asked, such as, is John friends with Mary?

```
In PROLOG: Does (John friends-with Mary)
Answer: YES
```

The advantages of PROLOG are evident from the fact that PROLOG has been chosen as a standard language for the Japanese Fifth Generation Computer Systems Project. In addition, it is also often claimed that the logical scheme is popular because of its very general expressive power and well defined semantics.

However, according to Genesereth and Ginsberg (1985) there are the following limitations to most logic programming systems:

- language constructs are very fine grained and do not provide adequate facilities for defining more complex constructs;
- generality of the predicate calculus is a barrier to the development of effective deduction facilities for using knowledge expressed in it.

On the basis of this short description of three main ways of representing knowledge it can be concluded that no single representation formalism seems adequate, and that each technique has its own strengths and weaknesses. To solve this problem, some hybridizations have also been developed.

This review has also shown that knowledge representation is a central issue in expert systems. This will be helpful in a discussion about the fundamental problems, relevant for the social sciences. To complete the picture, it is also necessary to say something about the development of expert systems, and, in this context to stress the differences between "classic" expert systems and so-called "shells" as additional example of disagreements between authors.

2.4. Some features of expert systems development: from classic expert systems to "shells"

The earliest development in expert systems began in the area of structural chemistry within the DENDRAL project (see Lindsay et al., 1980). Organic chemistry is one of the most appropriate fields for expert systems development because it,

> "... has a strong formalism, that of the graphical structure diagram or its network equivalent within the machine, to represent molecular structures, and transformations among them" (Town et al., 1985, Section 3.1., p.1).

The whole DENDRAL project includes three programs: DENDRAL, CONGEN, and META-DENDRAL.

The major program HEURISTIC DENDRAL was the first and is probably the best known expert system (e.g., it has recently been described as the grandfather of expert systems - see Aleksander, 1984). The program is designed for use by organic chemists to infer the molecular structure of complex organic compounds from their chemical formulas and mass spectrograms (mass spectrograms are essentially bar plots of fragment masses against the relative frequency of fragments at each mass). The program was developed by E. Feigenbaum, B. Buchanan and others in 1965 at Stanford University.

The program makes use of rules which relate physical features of the spectrogram (high peaks, absence of peaks, etc.) to the need for particular substructures to be present in or absent from the unknown chemical structure. Using these constraints, CONGEN (see Carhart, 1979) produces a list of all acceptable candidate structures. For each of these a spectrogram is then computed, and a matching algorithm ranks the candidates in order of the best fit between their spectrograms and the spectrogram of the unknown compound. While CONGEN resulted from the slow speed of DENDRAL, META-DENDRAL (see Buchanan and Feigenbaum, 1978) is an attempt in the automatic acquisition of knowledge.

Another pioneer system which has already been described in detail, is MYCIN. In this context, it is, perhaps, useful to say that its framework, EMYCIN, has led directly to SACON, an advisory program for structural analysis in engineering, and to PUFF which analyses results of pulmonary function tests for evidence of possible pulmonary function disorder. An offshoot of MYCIN is TERESIAS (see Davis et all, 1977) which concentrates on knowledge acquisition, i.e., it assists in the construction of large knowledge bases by helping transfer expertise from the human expert by means of a dialogue.

MYCIN's simple knowledge structure has stimulated other developments directly, where, among others, PROSPECTOR and XCON (earlier named R1) deserve a brief description.

PROSPECTOR is a system developed by R. Duda, P. Hart and others at SRI International in California. It is intended to aid geologists in assessing the favourability of a given region as a site for exploration for ore deposits of various types. The user provides the program with a list of rocks and minerals observed "in the field" and other information expressed in a rudimentary form of English. The program then conducts a "dialogue" with the user, requesting additional information where needed. At any point, the program is able to provide the user with an explanation of the intent of any question. The eventual output from the program is an indication of the "level of certainty" to which the available evidence supports the presence of a particular form of deposit in a given site.

The system has a number of different knowledge bases, corresponding to different classes of ore deposits. The program's knowledge for a particular ore deposit is held in the form of an "inference network" of relations between field evidence and geological hypotheses. The user expresses his certainty about a piece of evidence on a scale -5 to 5, where 5 denotes that the evidence is definitely present, -5 that it is definitely absent, and zero indicates no information. These are converted automatically into probability-like values. However, as it has already been said, PROSPECTOR provides a nice example of controversial claims about the achievements of expert systems.

XCON (earlier named R1), developed in the late 1970s, is a rule-based expert system that configures VAX computers (see McDermott, 1982). It takes the specifications for a new computer installation, determines the physical layout and interconnection of the computer's components, checks the resulting configuration for order and consistency, and, if necessary, either upgrades hardware specifications (introduces a heftier power supply) or adds missing components (a cable). According to Duda and Shortliffe (1983) it is now used by the Digital Equipment Corporation (DEC) to configure every VAX that is sold. It is claimed that it results in a $10 million annual saving; and 85% of configurations are reported to be faultless (see Hayes-Roth et al., 1983). However, there are two questionable issues:

- the system is much more difficult to amend than a straightforward program would be;

- when tested by the user community, the performance of the system declined to the 60% level (see Davis, 1984).

The interest in building expert systems has been widening significantly. The following are some of the well-known examples of such expert systems: INTERNIST (used for diagnosis in internal medicine), SECS (proposes schemes for synthesising stated organic compounds), DIPMETER (advises on oil-well driling), MOLGEN (assists in the design of experiments in molecular genetics), etc.

Of course, this list could be much more comprehensive; the main difficulty when producing such lists is in the lack of firm evaluation criteria, the most worrying being the lack of users' appreciations of existing expert systems.

The systems described above can also be defined as "classic" expert systems, recently being estimated as "distrusted" and "disused" (see Pogson, 1986). Such statements mainly criticize the building of expert systems from scratch which often takes at least five man-years of effort (see Davis, 1984). An alternative is to make use of a standard framework or "shell" which enables working systems to be developed rapidly. As a result, a number of commercially available "shells" have appeared in the last few years.

However, expert systems "shells" are again a topic of many disagreements. One the one hand, authors like Gooding (1986) claim that this is,

> "... the only route by which expert systems can make a contribution to mainstream computing" (p. 39).

On the other hand, D'Agapeyeff (1984) found in surveying the commercial applications in the UK that it was necessary to introduce a new term "simpler expert systems" to adequately categorise much activity. His conclusion about "shells" and their related commercially available expert systems is that,

> "It is necessary to correct the impression, much heralded hitherto, that Expert Systems are inherently complex, risky and demanding. This impression is a handicap to competetive developments in the supply and usage of Advanced Information Technology" (p. 3).

There is no doubt that using a "shell" can greatly reduce the development time of an expert system because all "shells" provide a following basic framework:

- a means of encoding the domain knowledge;
- inferencing mechanisms (typically backward chaining) for making use of the encoded knowledge.

Therefore, with the complex programming tasks being done by the "shell", the task of building an expert system is greatly simplified and the builder is free to concentrate on the knowledge acquisition process. This is a much quicker and cheaper process than building expert systems from scratch.

As we have already seen, the first and probably the best known "shell" is EMYCIN; the creation of PUFF and SACON has also been mentioned. According to Johnston (1986), dozens of "shells" working on this general principle are now on the market and are achieving some success.

In addition, there is a trend of developing expert system "shells" for microcomputers (see survey by Guilfoyle, 1986a). Amongst these the most popular are Xi (uses "if ... then" rules and forward and backward chaining) and Guru (integrates expert systems building tool with spreadsheet, database manager, text processor, etc.).

Despite the commercial success of expert system "shells", it is necessary to say a word of caution:

- the problem of an overall knowledge representation scheme and "backward chaining" has already been discussed in the context of EMYCIN - when applying "shells" it is always necessary to have in mind the differences between subject domains;

- expert systems "shells" are primarily commercially oriented, and their builders are not interested in some of the "academic" issues of AI, for example, the ability to learn, etc., which originally excited researchers in AI. Therefore, there is a danger of the transformation of expert systems into a "... flabby marketing phrase" (Gooding, 1986, p. 39). Or, in other words,

> "AI is a chaos. It's hard to get good researchers to work on fundamental problems because the companies are snapping them all up. Theory has stagnated for a moment, and we've lost our momentum" (Waldrop, 1984, p. 804).

At the end of this section it can be concluded that this discussion has so far revealed some important characteristics of expert systems research. One of the most alarming is the lack of agreement among authors about even the basic issues, such as definition, aims, evaluation, commercial use, etc., of expert systems. The additional problem is in over-simplified descriptions of the field and in its atheoretical foundations.

At this point the question can be asked: what is impeding greater clearness in expert systems discussions?

Is the reason only in the fact that this work is relatively new? I think not.

In this study, the central role of knowledge in building expert systems has often been repeated. Therefore, the answer to the above question can only be provided in a detailed analysis of crucial expert systems issues (which can also be defined as fundamental problems), such as knowledge acquisition and representation, and the connected explanation facility. According to their nature, there is no doubt that these issues should also be addressed to the social sciences.

2.5. Fundamental problems in expert systems research, relevant to the social sciences

2.5.1. Knowledge acquisition

Knowledge acquisition for expert systems is a difficult and time-consuming process. Barr and Feigenbaum (1981) describe it as the biggest bottleneck in the production of these systems.

Knowledge acquisition is defined as,

> "... the transfer and transformation of problem-solving expertise from some knowledge source to a program. Potential sources of knowledge include human experts, textbooks, data bases, and one's own experience" (Buchanan et al., 1983, p. 128).

It is claimed by Hayes-Roth et al. (1983) that this transformation is the heart of the expert system development process. In this context, the role of the knowledge engineer is introduced whose function is to extract the knowledge from the relevant expert and to code this knowledge for input to the computer. Through an extended series of interactions, the knowledge engineering team (the knowledge engineer and the expert) defines the problem to be attacked, discovers the basic concepts involved, and develops rules that express the relationships existing between concepts. Thus, there are the following major stages in the evolution of an expert system, as identified by Hayes-Roth et al. (1983) and shown in Table 3.:

Identification:	Determining problem characteristics
Conceptualization:	Finding concepts to represent knowledge
Formalization:	Designing structures to organize knowledge
Implementation:	Formulating rules that embody knowledge
Testing:	Validating rules that embody knowledge

Table 3. Stages in the evolution of an expert systems (after Hayes-Roth et al., 1983, p. 24).

However, the lack of emphasis placed on the techniques (or problems) of extracting expert knowledge and converting it into a suitable form (generally rules) in the "popular" literature on expert systems might lead to a conclusion that it presents no difficulties.

Unfortunately, very little is known at present about how to extract expertise from an expert and almost nothing is on offer as a technique. In other words, as discussed in a report by Welbank (1983) who examined a variety of knowledge acquisition techniques for expert systems and concluded that this field is,

> "... at a very early stage of development, where different experiences are still being gathered, and general principles have not emerged" (cf. Bramer, 1985, p. 7).

The problem of knowledge acquisition is characterized on the one hand by an expert, unfamiliar with expert systems and unable to articulate what knowledge he has and how he uses it to solve problems; and on the other hand by a knowledge engineer who may well be totally ignorant about the domain of expertise.

In addition, the most effective methods of acquiring knowledge from experts, such as observation "in the field" or in-depth interviewing, are inherently slow, a major problem given that experts' time is often in short supply. While an experienced team can put together a small prototype system in 1 or 2 man-months, the effort required to produce a system that is ready for serious evaluation is more often measured in man-years. There is no doubt that these methods are often expensive and also prone to error.

Some have argued that the best way to overcome the problems associated with traditional techniques of knowledge acquisition is to move towards automatic methods of rule generation based on an analysis of example cases. The essential idea behind rule-induction - based on

Quinlan's ID3 algorithm (see Quinlan, 1979) - is that given a database of examples, machine induction can quickly generate a rule base which completely accounts for all the examples, and in general this can be performed in many different ways. The use of this approach is nicely described by Michie (1984) who claims that expert systems can now also be used,

> "... to put the knowledge back into human hands in improved form" (p. 342).

One of the successful examples of using this method can be found in a study by Mozetic, Bratko, and Lavrac (1983). Using the logic programming language PROLOG, the authors collaborated with senior clinical cardiologists at the Ljubljana University Medical School. The Yugoslav group constructed a computer model of a complete and ultra-reliable diagnostic scheme for multiple arrhythmias and their relation to the ECG wave form. The system produced new knowledge, although small in extent, but sufficient to have a use in teaching and as a reference text for the specialist.

However, it has also been claimed that automatic rule induction has some weaknesses (see Bramer, 1985), for example:

- automatic induction may result in a set of rules that is formally correct (in the sense of accounting for all the examples given) but which has low predictive power for the cases outside the example set;

- the possibility of "noise" in data values and the possibility that some necessary attributes (perhaps those which are only significant for a fairly small number of cases) are missing.

Such reservations should not reduce the role of rule induction methods which are at present one of the essential steps in the development of expert systems. Another, more general, problem connected with knowledge acquisition techniques is the question of whether an expert's knowledge can be represented in its entirety in a computer program.

2.5.2. Knowledge representation

It is claimed in the literature that specialized knowledge encapsulated in expert systems is of two types:

> "The first type is the facts of the domain - the widely shared knowledge ... that is written in textbooks and journals of the field, or that forms the basis of a professor's lectures in a classroom. Equally important to the practice of the field is the second type of knowledge called heuristic knowledge, which is the knowledge of good practice and good judgment in a field. It is experiental knowledge, the "art of good guessing", that a human expert aquires over years of work" (Feigenbaum and McCorduck, 1984, p. 76-77).

In addition, knowledge engineers are supposed to know how to extract relevant knowledge from an expert and how to encode that knowledge in a form amenable to mechanical manipulation.

As it is evident from the above quotation, it is thought that the essential idea behind expert systems building is that all human knowledge can be fully exhausted by facts and heuristics.

A word of caution about such claims can be found by Aleksander (1984) who emphasizes that,

> "... there are .... many more areas ... where knowledge cannot be encompassed in simple logical or probablistic rules" (p. 134).

An alternative and very useful view on knowledge elicitation is explained by Collins, Green and Draper (1985) who introduce the importance of so-called "cultural, tacit, and skilfull aspects of knowledge". Their main hypothesis is that when domain expert's knowledge is elicited and encoded, these aspects of knowledge are lost. This is explained by a very illustrative metaphor: knowledge is like chicken soup with dumplings, and the expert system is like a colander; with all known expert systems, the dupmlings get transferred but the soup is lost. The dumplings are the readily explicable facets of knowledge such as factual information and articulateable heuristics, whereas the soup is the context/meaning of the facts and the non-articulated but "taken for granted" practices and "ways of going on" in practical and theoretical settings, or in other words "tacit knowledge"(17). This has implications for the use of expert systems, i.e., expert systems must rely on the users' abilities and their "tacit" knowledge to interpret the system's advice. Consequently, it is clear that the more expert the end-user, the easier it will be to build a system that will be useful.

These authors conclude that the crucial division in knowledge is not the separation between facts and heuristics, as much work on knowledge elicitation has stressed, but between the articulateable and the tacit. The promise and development of expert systems can be much better understood once it is realized that limits are set by the fact that substantial components of knowledge are not articulateable.

But, I believe it is necessary to be much more precise when making this division in knowledge. Additional distinctions should be made between domains where knowledge can be represented in a highly structured, formalized way (e.g., different domains in chemistry, mathematics, etc., where we can find immediate contact with common sense, or so-called "tacit" knowledge) and the areas where knowledge cannot be represented in strong formalism (e.g., different areas in the social sciences). Expert systems can be, of course, much more reliable in the former area. It is surprising how infrequently this problem is discussed in the literature on expert systems. Much of the descriptions and analysis are based on the assumption that defining "narrow domain" is the ultimate condition for building expert systems, without taking into account the characteristics of the knowledge structure of this domain. Starting from the structure of knowledge, many confusions and disagreements about expert systems can be avoided.

This leads to a conclusion that developments in expert systems depend not only upon advances in knowledge engineering, but also on research in the wider fields of AI and the social sciences which underpin the complex approach to knowledge, i.e., transfer of knowledge, formalization of knowledge, common sense, etc.

2.5.3. Explanation facilities

It has been stressed in this study that the explanation facility of most expert systems consists of nothing more than printing out a trace of the rules being used. These facilities are valuable, not least in the possibility that the end user can learn about the knowledge domain by interacting with the explanation facility (e.g., GUIDON).

However, many expert system explanation facilities cannot fulfil a much more important role. Michie and Johnston (1985) put the matter this way:

> "Any socially responsible design for a machine must make sure that its decisions are not only scrutable but refutable. That way the tyranny of machines can be avoided" (p. 69).

In this context, Michie and Johnston talk about a "human window" (as opposite to a lunatic black box) in computer programming - a window of reasoning that is like human reasoning in depth and complexity.

At this point, the question can be raised as to whether the type of explanations described above can make a machine's decisions refutable. The basis of these explanations are rules and encapsulated knowledge. Therefore, when discussing the reliability of explanation facilities, it is very important to start again from the understanding of structure of knowledge and its formalization in a particular applied domain. There is no doubt that in domains where knowledge cannot be represented in a highly formalized way, these explanations are only condensed fragments of the expert's knowledge. There is a whole host of "tacit" knowledge (see Collins, 1986) which expert systems cannot handle but which is essential for the provision of a good explanation: that is, an explanation that can be refuted.

Refutability is important because explanations are not simply extras which are provided by expert systems. D. Michie has long been concerned with the refutability of computer programs both from the point of view of producing a good system, and also because non-refutable systems can cause catastrophes when used in such areas as air traffic control, air defence or nuclear power (see Michie and Johnston, 1985).

The doubts about whether current expert systems can provide refutable explanation facilities are expressed by Leith (1986), who claims:

> "Unfortunately, it is beginning to seem as though expert systems have not been designed in a socially responsible way, for they cannot really explain the basis for their reasoning in as full a manner as the non-expert needs" (p. 15).

Collins (1986) goes even one step further in his criticism and argues that current expert system explanation facilities can do nothing but add to the friendliness, persuasiveness and seeming authority of the formalized knowledge encoded within an expert system. He proposes that an explanation facility which cannot be refuted should be banned,

> "... and this means cutting out the explanation facility except where the expert system is to be used by an expert!" (p. 9).

I think that it is important to repeat here how little has been said in these discussions about the structure of knowledge in a particular domain.

The fact is that the expert systems community is, nevertheless, aware of the inadequacy of current expert systems explanations (see Guilfoyle, 1986b). The following are some attempts to improve the explanation facility which indicate the importance of this component of expert systems:

- use of additional knowledge beyond the system's performance rules;

- user modelling: building up a picture of the user, which can help to tailor output, interfaces, help levels, and so on, to the particular user; this is an assessment of what the user does and does not know and what he is trying to accomplish;

- increased use of diagrams which can sometimes offer better explanations than text, etc.

At the end of this chapter it can be concluded that there is no doubt that three fundamental issues in expert systems (i.e., knowledge acquisition, knowledge representation, and explanation facility) are also relevant concepts in the social sciences and are, as such, legitimate subjects of social sciences research. This idea is very important: On the one hand, expert systems will be able to meet the challenge of general competence and reliability if more fundamental progress is made by AI, psychology, sociology of knowledge, etc., in understanding the structure of knowledge and the whole knowledge complex (e.g., transfer of knowledge, formalization of knowledge, process of reasoning, common sense, etc.). On the other hand, this understanding can be a starting-point for seeing where the development of expert systems is going and how it will get there. Only such an established framework will enable the social sciences to discuss realistically the problems of the impact and effects of expert systems and AI.

The role and potential of AI and expert systems for the library/information community can also be understood from this point of view. This question was raised in the introductory section and will be explained in the following chapter.

Chapter 3

The relevance of AI and expert systems research for library/information systems

> "The goal of expert systems research is to provide tools that exploit new ways to encode and use knowledge to solve problems, not to duplicate intelligent human behaviour in all its aspects" (Duda and Shortliffe, 1983, p. 266).

Throughout this discussion it has been emphasized that knowledge representation is one of the fundamental problems in expert systems building. However, without regard to the difficulties in trying to encode expert's knowledge in representation schemes, it should be stressed that research on knowledge representation methods and techniques is one of the vital issues in the whole field of AI. The results of such research can be important to a variety of scientific and economic endeavours, including the design of improved library/information systems.

At present there are two main identified areas where the achievements of AI and expert systems research can be useful for libraries and information services, as follows:

- expert systems building and knowledge representation schemes force a rethink of the methods of organizing and representing information and knowledge in databases in order to make it dynamic and interactive;

- the development of expert intermediary systems as front ends to bibliographic databases.

3.1. A need for new methods of organizing and representing information and knowledge in databases

It has already been said that information science which grew out of the library science with the introduction of computers and which is,

> "... concerned with formalizing the process of knowledge formulation, organization, codification, retrieval, dissemination, and acquisition" (Walker, 1981, p. 348),

and artificial intelligence are relatively new areas of research, each having assumed an independent identity within the past thirty years.

The relevance of much of the research in AI to library/information systems seems to be in the middle step in the "information transfer cycle", i.e., computerized information storage and retrieval.

But first, to provide a theoretical framework for a discussion it is necessary to clarify the term "database systems".

According to Town et al. (1985) a database system may be considered as consisting of three major components. Firstly, there are the records that form the body of the database, each record comprising one or more data elements. These elements may be of several types, for example:

- numeric data;
- structural data;
- textual data.

Secondly, there are the search mechanisms that allow a user to query the database so as to retrieve records from it. Thirdly, there is the interface to the database by means of which a user may specify his or her query.

Therefore, a distinction can be made between database systems which provide the user with direct access to source data, either a variety of different kinds of numerical values (also numerical databanks), structural data (e.g., structure-based systems in chemistry), or even the complete texts of documents (in the legal area, e.g., LEXIS) and, in contrast, bibliographic databases, which help the user to identify primary or source documents that might have information relevant to his needs and interests.

Although it is important to develop so-called "system thinking", as defined by Kornhauser (1983), i.e.,

> "... an organized way of linking bits of information into networks, trees, modular systems, showing the interrelationship between data" (p. 385),

in all three kinds of databases, I would like to concentrate, in this section, on the possible usefulness of knowledge representation research to rethinking the organization of information in bibliographic databases. The problems of interfaces to database systems will be outlined in the next section.

One of the main disadvantages of bibliographic systems is that they only provide pointers to literature; the user can make a preliminary assessment of the utility of a reference from the title and abstract, but he still has to find the documents and evaluate their contents before he can derive information from them. This is not a straightforward matter like looking up an item in a table in a databank.

The use of bibliographic databases also relates to the procedures for classifying or indexing the document, i.e. the inclusion of a document in a database requires that a judgment be made about its content. But, however appropriate index terms or thesaurus entries to a document (either manual or automatic indexing) are, there is one main drawback: index terms are words, and they can have in isolation many meanings. Consequently, document searching often produces much irrelevant material.

The fact is that, if one wanted to create a model with respect to user requests, this model,

> "... needs to be more than just a list of index terms, but to be terms in relationships" (Addis, 1982, p. 302).

In this context, the most acceptable structures are those that best maintain the semantic feature of information. Currently, this requirement is best fulfilled by the previously described knowledge representation scheme, called semantic nets.

This is one of the possible methods which can be used as a complement to the subject oriented approach of most abstracting and indexing services. Hjerppe (1983) identifies within this notion three different applications, i.e.:

- condensing existing knowledge in stages;

- organizing existing knowledge to show structure and relations of documents;

- organizing existing knowledge to exhibit lacunae and unnoticed links.

An example of an attempt at the first application mentioned is the Hepatitis Knowledge Base (HKB) which was created in the USA for medical researchers and doctors interested in hepatitis (see Bernstein, Siegel and Goldstein, 1980). This database which can be searched on-line, comprises information initially extracted from forty review articles and then expressed in a series of hierarchical statements to form a consensus of all the available knowledge about the disease. The problem of knowledge acquisition and updating for such a database is described by Walker (1981).

Examples of the second and third applications mentioned can be found in the co-citation clustering concept. According to Small (1986), a co-citation cluster is a bibliometrically defined network structure, and the hypothesis is that it defines a knowledge structure as well as an "invisible college" or social structure. In essence, the method pieces together selected passages from a variety of sources to form, as far as possible, a coherent whole. The sources of text for the narrative are the papers which cite the core documents in the cluster, and specifically the context of citation for those documents. Using this method it has also been shown how paragraphs which cite multiple core documents in a cluster can be used to provide an interpretation of the structure of the co-citation network map (see Small, 1984). Earlier works on the database ISI/BIOMED and recent work in the field of education (see Ward and Reed, 1983) indicates the need for new approaches in the organization and representation of information and knowledge in databases.

In addition, there are also a number of interesting attempts towards building expert systems for "traditional" library work, such as cataloguing (see Davies and James, 1984; Hjerppe, Olander, and Marklund, 1985), reference services (see Bivins and Palmer, 1981) and some other important library/information tasks.

3.2. Expert intermediary systems as front ends to bibliographic databases

It has been emphasized in the previous section that the interface to databases is also one of the essential elements in the whole database system, particularly because,

> "The flexibility and ease of use of this interface will play a large part in determining the degree to which the database is utilised by the intended user community: systems that require extensive effort to express a query, or which provide little feedback during the retrieval operations, are unlikely to be used at all heavily" (Town et al., 1985, Section 2.1., p. 1).

Therefore, a flexible and user-friendly interface is needed. There are two approaches which derive from AI research and could be appropriate to such a type of interface:

- question answering systems (e.g., LUNAR, LADDER, etc.);

- expert systems interfaces, i.e., their explanation facilities (e.g., MYCIN).

Both approaches have already been described in this study and it has been stressed that their main advantage is that they allow communication with a database via the user's natural language. Their main drawback is domain dependence. Another important type of interface, not discussed here, is the use of graphical interfaces (e.g., applying graphics techniques to numeric and textual files; see Michard, 1982).

When desribing interfaces it is also necessary to stress the difficulties connected with the problems of access to on-line bibliographic databases. According to Town et al. (1985), there are currently over 500 online bibliographic databases containing more than 100 million citations, while a full text file may contain several tens of billions of characters. But, at the present time many users are unable to make full use of available computer searchable databases as these require knowledge of both the mechanisms of performing a search and of the way a controlled vocabulary may be used to express a document search request.

Because of these difficulties, online searches are usually carried out by professional intermediaries instead of end-users. Reference librarians and intermediares are needed to help formulate user requests in terms of the information systems and to provide guidance on how the system is organized, on what materials are available, and on how to search for and locate the desired items. This is a very questionable situation because it is very difficult to determine what the user really requires and the searcher may seldom be aware of his own real needs. The connection is, in Pollitt's words (1986), that computerized searching services will not have their full impact upon user communities until direct user searching is widespread.

Of course, there has been considerable interest, especially as the result of AI research, in how to make request formulation easier and more effective. Of greatest potential here are expert intermediaries that can function as front ends to an existing searching system and which enable the user to undertake good quality searches without the knowledge or training demanded of the professional search intermediary.

Given the complexity of the problem domain of online searching, it is evident that considerable expertise is reqired to enable good decisions to be made and the search to be conducted. There is not yet a detailed taxonomy of this expertise, although various writers on the design of front end systems have suggested categories which could be used to characterize this expertise. For example, Pollitt (1981) lists four categories:

1 - system knowledge: the command language and facilities available in the search system(s) from logging on and the submission of search statements to the printing of references or abstracts;

2 - searching knowledge: relating to the strategy and tactics to be employed in searching;

3 - subject knowledge;

4 - user knowledge: knowledge about each individual user including previous searches and preferred journals.

It is added by Smith (1986):

5 - database knowledge: familiarity with the content and structure of available databases.

The fact is that more research is needed to determine in more detail the expertise underlying successful searching.

A general review of expert intermediary systems is given by Marcus (1983). The main characteristic of these systems is that users are freed from encounters with the many peculiarities in databases and search systems, and yet can benefit from a large range of capabilities, for example:

- users can enter a request in a loosely structured format, preferably in a natural language, sentence-like expression. An intermediary system processes the request terms, displays information to the user (in the form of a list of subject areas, databases, search keys) from which users are asked to make a selection. Interactions of this nature usually proceed until users terminate the session;

- intermediary expert systems can also replicate the performance of an expert in a particular area by incorporating the knowledge of an expert with rules for making inferences on the basis of this knowledge.

The most interesting examples of front end systems are CONIT, EXPERT-1, and CANSEARCH which all are being tested in experimental settings.

CONIT (see Marcus, 1983) is a system that allows end users, who had no previous experience in operating retrieval systems, to obtain information, i.e., literature citations, from dozens of heterogeneous databases on four different host computer systems. This system emphasizes a command/argument language structure to the interface. A development of CONIT, called EXPERT-1 seeks to simulate a human expert's search procedures in terms of search strategy formulation and explanation. The expertise built into EXPERT-1 includes the following abilities:

- to assist the user to formulate his search problem;

- to formulate a search strategy based on the concepts and search terms supplied by the user;

- to assist the user to select the appropriate databases to search in;

- to handle different protocols and command languages;

- to assist the user reformulate search strategy based on partial search results.

While CONIT - and most of the other intermediary systems - emphasizes a command/argument language approach to the user interface, EXPERT-1 employs only menus and the fill-in-the-blank mode of computer - human interaction.

A similar systems, CANSEARCH has been described by Pollitt (1984; see also 1986) to help doctors carry out online searches of cancer therapy in the MEDLINE database. Here again, the use of a human intermediary is replaced by a series of hierarchically organized menus that guide the user through the process of identifying those components of their problem that may need to be included in a query formulation. Thus, menus are available for the specification of the site of the cancer, its type, the therapy that is under consideration, and the characteristics of the patient. Doctors can select the appropriate part of the menu on a touch-sensitive screen. When all of the components of the search have been identified to the doctor's satisfaction, the system generates a Boolean query that is then submitted for processing by the MEDLINE system. The components of the

CANSEARCH system are shown in Fig. 8.:

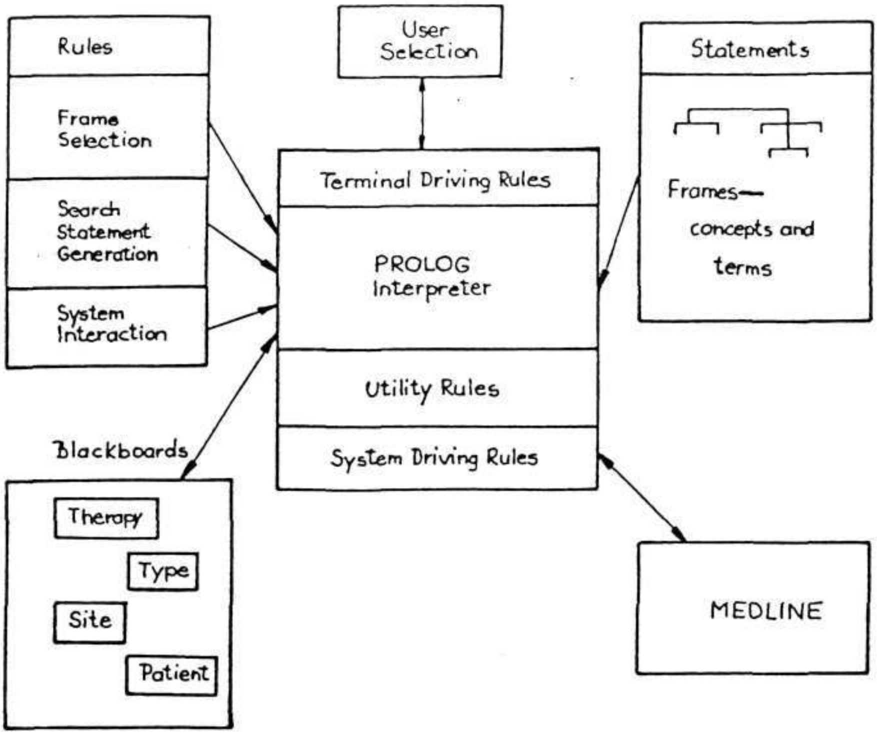

Fig. 8. Components of the CANSEARCH system (after Pollitt, 1984, p. 233).

According to Town et al. (1985), such a design methodology would seem appropriate for end-user access to any reference database for which a well-designed thesaurus is available and for which the domain of possible query types was similarly restricted. It is also interesting to stress, according to the same source, that the Commission of the European Communities has recently examined proposals for a DIANE Intelligent Interface Facility

(IIF). The aim of IIF is to improve access to the databases available on the Euronet DIANE telecommunication network and, as far as possible, to make the network transparent to the user.

Deriving from the relevance AI and expert systems research in designing improved library/information systems, it is in the library and information profession's interest to be aware of these developments. There is no doubt that the role of this profession is changing, for example, illustrated by Clarke and Cronin's statement (1983):

> "On the basis of the results of early trials there is a strong indication that the user of the future will not need the services of a librarian or information scientist in order to be able to conduct a comprehensive and successful on-line literature search" (p. 286).

Consequently, this means that the changes resulting from the application of AI research in library/information services must also have an impact on library/information science education, i.e., it is the task of library/information schools to prepare students for new roles and new careers in information transfer.

CONCLUSIONS

Throughout this discussion, the explosive growth of interest and work in AI and expert systems has been illustrated which, after optimistic beginning in the late 1950's and a period of stagnation in the early 1970's, began again in the early 1980's. This is evident from the Japanese Fifth Generation Computer Systems Project, the Alvey Programme in the U.K., ESPRIT, etc. One of the characteristics of this growing interest is that, on the one hand, many governments in the West provided resources for research, and, on the other hand, commerce and industry showed much greater openness to the ideas, techniques and tools of AI. The reason for this interest is, of course, the fact that AI is becoming a big, and increasingly growing business, most notably expressed in the areas of expert systems, natural languages, vision and robotics.

Today, there is no doubt that AI has captured media interest and is fast becoming one of the most important topics in discussions about the development of new technology. The indicators of this are the numerous articles in newspapers and journals, recently published books about this topic, newly established journals devoted only to this subject, etc. These published accounts relate to different disciplines, which emphasizes the position of AI as an interdisciplinary field. The reasons for the multidisciplinarity of AI lie, of course, in the term "intelligence" which is not well-defined and is, as such, included in many disciplines.

In the context of the increasing interest in AI research, especially with regard to its commercial implications, there is one worrying feature: there are wide discrepancies in the reports about the achievements in the field. Throughout this study, many examples have been quoted, one of the most alarming being PROSPECTOR, an expert system the aim of which is to aid geologists in evaluating mineral sites for potential deposits. On the one hand, this expert system is the subject of many claims about the savings it had made for exploration companies, and, on the other hand, the designers of PROSPECTOR are impugning such statements as untruth.

Variations in the assessments of the state of the art were the starting-point for establishing a need for a closer association between AI and the social sciences, until the present only used by AI researchers to investigate the effects and impact of AI. If the reasons for disagreements about AI are to be understood, a social sciences approach - with sociology taking the central role - to the discourse of AI has to be applied.

It has been stresed in one sections that the reasons for discrepancies in reports about AI and expert systems lie in the interpretative flexibility of the term "intelligence". What counts as "intelligence" and the questions about whether machines can be intelligent or not are at the heart of these discrepancies. For example, optimistic views, representations of the achievements of the field might be expected from those involved in marketing AI applications. However, these reports are elsewhere (in the less "popular" press) countered by considerable caution and pesisimism about the achievements to date. To recognize the interpretative flexibility of this idea of "intelligence" is a very important goal for the social sciences because otherwise they will have to wait for the output of AI research, rather than be involved in a detailed consideration of the process of the research activity itself.

It follows that the social sciences should not only be concerned with the effects of AI, but also with its genesis. The social sciences will only be capable of assessing realistically the impact and effects of AI research if they also have insight into the key issues of AI, one of the most important being knowledge representation. There is no doubt that knowledge is also a social concept, which again indicates the need for a closer relationship between AI research (e.g., knowledge engineering) and the social sciences.

For many years it has been thought that the development of cognitive psychology and the so-called "computational metaphor" is the only relationship between social sciences and AI. In addition, many introductory sections of AI literature also include citations about the relevance of psychology and linguistics to AI. What is surprising is the fact that the important possibility of an association between sociology and AI has hardly been noticed. There is no doubt that sociology, whose interests clearly encompass language use and interaction, intentionality, knowledge systems, etc., is as yet unexplored territory. And, perhaps even more important, knowledge is recognized in the modern sociology of knowledge as a legitimate sociological object. Therefore, a two-sided exclusion of sociology from AI research has been discussed:

1 - AI researchers have not been interested in the possible contributions of sociological research in, for example, knowledge systems;

2 - the function of sociology has been reduced to the investigations about the impact and effects of AI, instead on the AI research activity itself as a condition for such investigations.

It has been shown that this reduction of sociological capability corresponds to the pre-Kuhnian view of science, where "social factors" were precisely those factors not related to "science itself"; the domain of the "social" was regarded as outside or (at best) peripheral to the actual science. But as the post-Kuhnian sociology has established the nature and content of scientific knowledge as legitimate object, the task of sociology is to break the barrier between "the social" and "the scientific" in the context of AI research.

In the second part of this study, the assumption that the social sciences should also be concerned with the genesis of AI was tested by the example of expert systems.

Although expert systems are recognized as the applied end of AI research, there are many controversies in the field. Extraordinary optimism is very often countered with reports that the area faces fundamental problems. An additional problem is a lack of agreement in expert systems literature about even the basic issues such as definition, aims, essential characteristics, evaluation, commercial use, etc., of expert systems. Moreover, many over-simplified descriptions of the field can also be found. Such an "atheoretical" foundation leads, of course, to difficulties when trying to assess the future development of expert systems, especially with regard to their their possible effects and impact in different environments, issues which are usually addressed to the social sciences.

From this starting-point the question as to what was impeding greater clearness in expert systems discussions was raised. It has been stressed that only a detailed analysis of crucial expert systems issues, such as knowledge acquisition, knowledge representation, and explanation facilities can provide the answer to this question. There is no doubt that these issues are also social concepts and, as such, relevant for the social sciences. This notion is very important because, on the one hand, expert systems will be able to meet the challenge of general competence and reliability if more fundamental progress can be made by the social sciences (e.g., psychology, sociology of knowledge, etc.) in understanding the structure of knowledge and the whole knowledge complex, i.e., transfer of knowledge, formalization of knowledge, process of reasoning, common sense, etc.; and, on the other hand, this understanding is a starting-point to see where the development of expert systems is leading. Only such an established framework will enable the social sciences to discuss realistically problems of the impact and effects of expert systems (e.g., how the status of the human experts will be affected by expert systems, legal implications of the use of expert systems, etc.).

Throughout this study the notion that knowledge representation is a vital issue in AI research today has been followed, i.e., its goal is to provide tools that exploit new ways to encode and use knowledge to solve problems. Without regard to some fundamental problems connected with knowledge representation schemes, such research results can be important in many different environments, including the design of <u>improved library/information systems</u>. Two areas where such research can be useful for library/information services have been identified:

- providing new methods for organizing and representing information in databases, i.e., "system thinking" which means an organized way of linking bits of information into networks, showing the relationships between data;

- developing expert intermediary systems as front ends to bibliographic databases.

In this connection it has also been emphasized that these changes resulting from AI and expert systems research should have an impact on library/information education.

Finally, it is appropriate once more to stress that when applying new methods of organizing information and knowledge in databases it is necessary to put the structure of knowledge in the first place, i.e., it is necessary to recognize a distinction between fields where knowledge can be formalized in a highly structured way (chemistry, mathematics) and fields with "weak" formalism (social sciences). In this light - apart from the unjustified distinction between information services and libraries - the research project "Development of scientific and technical information in Slovenia 1986-90", mentioned in the introductory section, should be considered. Its relevance is in stressing the need for the development of new methods of organizing information and knowledge in database systems, but in doing that it should also take into account different relationships between knowledge, communication, and information systems in the sciences, the social sciences, and the humanties.

NOTES

1. These differences can be illustrated by the example of chemical information systems where the techniques used fall into two categories: firstly, those which involve the analysis and organization of text; secondly, those which are concerned with handling chemical structural information. Chemical data is unusual because the second aspect of it, the structure of a molecule, cannot be handled by normal bibliographic methods. For more information on this topic see Ash and Hyde (1975).

2. According to Becker (1986), Eugene Charniak and Drew McDermott trace in their book "Introduction to Artificial Intelligence" the first use of the terminology. They explain that in 1956 John McCarthy, an assistant professor of mathematics at Dartmouth College, and Marvin Minsky from MIT organised a conference in Dartmouth College in New Hampshire. During the conference McCarthy proposed that a study of AI should be carried out at the college to proceed on the basis of the conjecture that every aspect of learning, or any other feature of intelligence, can in principle be so precisely described that a machine can be made to simulate it.

3. Fundamental aspects of AI are clearly presented in books by Nilsson (1982), and Barr and Feigenbaum (1981).

4. Different positions to AI research are concisely described by Fleck (1984).

5. More about the construct "intentionality" can be found in Searle (1983).

6. This conversation is taken from Weizenbaum (1985).

7. Ed Weizenbaum (1985) illustrated the dangers of work in AI by one very interesting comment on ELIZA, written by an enthusiastic psychotherapist: "Further work most be done before the program will be ready for clinical use. If the methods proves beneficial, then it would provide a therapeutic tool which can be made

widely available to mental hospitals and psychiatric centres suffering a shortage of therapists. Because of the time-sharing capabilities of modern future computers, several hundred patients on hour could be handled by a computer system designed for this purpose. The human therapist, involved in the design and operation of this system, would not be replaced, but would become a much more efficient man since his efforts would no longer be limited to the one-to-one patient-therapist ratio as now exists" (p. 5).

8. LISP (short for LISt Processing) was developed by J. McCarthy and his associates at MIT during the late 1950s and early 1960s.

9. The programme is named after Mr. John Alvey of British Telecom, chairman of the committee which in 1982 recommended that such a national programme should be mounted, in response to increasing overseas competition and in particular to the Japanese Fifth Generation Computer Systems initiative.

10. Some details about the Alvey Programme are taken from its Annual Report 1985.

11. For example, the "Univision" system used by Unimation based on a vision system developed from AI research for the market by MIC, Machine Intelligence Corporation (see Winston and Prendergast, 1984).

12. The first report on PROLOG (PROgramming in LOGic) was first published in 1975 by researchers based at the University of Marseilles.

13. The examples of the dialogue with MYCIN are taken from Davis (1984).

14. There are also two alternative methods, i.e., propagation of constraints and problem reduction which have already been described in the section on the subareas of AI.

15. Knowledge representation schemes are succinctly described by Mylopoulos and Levesque (1983), and in three articles, edited by Friedland (1985).

16. Examples for PROLOG are taken from Michie and Johnston (1985).

17. The idea of "tacit" knowledge is taken from M. Polany (1976). Polany's well known example of tacit knowledge is the skill associated with bicycle riding. The formal dynamics of balance on a bicycle riding do not comprise the rules of riding. A rider may know nothing of centres of gravity and gyroscopic forces yet still rides whereas the most expert bicycle engineers may not be able to do so. The rider knows how to ride without being able to say how.

BIBLIOGRAPHY

Adams, J.B. (1976) 'A probability model of medical reasoning and the MYCIN model.' Mathematical Biosciences, 32, 177-186.

Addis, T.R. (1982) 'Expert systems: an evolution in information retrieval.' Information Technology: Research and Development, 1, 301-324.

Adler, F.R. (1984) An investment opportunity? In: Winston, P.H., and Prendergast, K.A. (eds.) The AI business: the commercial uses of artificial intelligence. Cambridge, Mass : MIT, pp. 255-262.

Aleksander, I. (1984) Designing intelligent systems: an introduction. London : Kogan Page.

Alexander, T. (1982) 'Practical uses for a "useless" science.' Fortune, 105, 138-145.

Alvey Programme, Annual Report 1985. London : Alvey Directorate.

Ash, J.E., Hayde, E. (eds.) (1975) Chemical information systems. Chichester : Ellis Horwood.

Barr, A., Feigenbaum, E.A. (eds.) (1981) The handbook of artificial intelligence. Vol. 1. London : Pitman.

Bateman, J. (1985) The role of language in the maintenance of intersubjectivity: a computational investigation. in: Gilbert, G.N., and Heath, C. (eds.) Social action and artificial intelligence. Aldershot : Gower Press, pp. 40-81.

Becker, J. (1986) 'On good with words.' Expert Systems User, 1, 20-21.

Belkin, N.J., Vickery, A. (1985) Interaction in information systems: a review of research from document retrieval to knowledge-based systems. London : British Library.

Bernstein, L.M., Siegel, E.R., Goldstein, C.M. (1980) 'The hepatitis knowledge base: a prototype information transfer system.' *Annals of Internal Medicine*, *43*, 165-222.

Bivins, K.T., Palmer, R.C. (1981) 'A microcomputer alternative for information handling: REFLES.' *Information Processing and Management*, *17*, 93-101.

Bobrow, D.G., Winograd, T. (1977) 'An overview of KRL, a knowledge representation language.' *Cognitive Science*, *1*, 3-46.

Boden, M. (1977) *Artificial intelligence and natural man*. Hassocks : Harvester Press.

Boden, M. (1984) 'Artificial intelligence and social forecasting.' *Journal of Mathematical Sociology*, *9*, 341-356.

Borko, H. (1985) 'Artificial intelligence and expert systems research and their possible impact on information science education.' *Education for Information*, *3*, 103-114.

Bramer, M.A. (1981) A survey and critical review of expert systems research. *In*: Parslow, R.D. (ed.) *Information Technology in the Eighties*. London : Heyden.

Bramer, M.A. (1985) Expert systems: the vision and the reality. *In*: Bramer, M.A. (ed.) *Research and development in expert systems*. Cambridge : Cambridge University Press, pp. 1-12.

Buchanan, B.G., Duda, R.O. (1983) 'Principles of rule-based systems.' *Advances in Computers*, *22*, 163-216.

Buchanan, B.G., Feigenbaum, E.A. (1978) 'DENDRAL and META-DENDRAL: their applications dimensions.' *Artificial Intelligence*, *11*, 5-24.

Buchanan et al. (1983) Constructing an expert system. *In*: Hayes-Roth, F., Waterman, D., Lenat, D.B. (eds.) *Building expert systems*. Reading, MA : Addison-Wesley, pp. 127-167.

Carhart, R.E. (1979) CONGEN: an expert system aiding the structural chemist. In: Michie, D. (ed.) Expert systems in the micro-electronic age. Edinburgh : Edinburgh University Press, pp. 65-82.

Cendrowska, J., Bramer, M.A. (1984) 'A rational reconstruction of the MYCIN consultation system.' International Journal of Man-Machine Studies, 20, 229-317.

Cercone, N., McCalla, G. (1984) 'Artificial intelligence: underlying assumptions and basic objectives.' Journal of the American Society for Information Science, 35, 280-290.

Clarke, A., Cronin, B. (1983) 'Expert systems and library/information work.' Journal of Librarianship, 15, 277-292.

Clocksin, W.F., Mellish, C.S. (1981) Programming in PROLOG, New York : Springer-Verlag.

Collins, H.M. (1986) The concept of explanation in expert systems. (Paper presented at conference on "Explanation in Expert Systems", University of Surrey, March 20-21, 1986).

Collins, H.M., Green, R.H., Draper, R.C. (1985) Where's the expertise?: expert systems as a medium of knowledge transfer. In: Merry, M.J. (ed.) Expert systems 85. Cambridge : Cambridge University Press, pp. 323-334.

Coulter, J. (1985) On comprehension and "mental representation". In: Gilbert, G.N., Heath, C. (eds.) Social action and artificial intelligence. Aldershot : Gower Press, pp. 8-23.

D'Agapeyeff, A. (1984) Report to the Alvey Directorate on a short survey of expert systems in UK business. London : Alvey Directorate.

Davis, R. (1982) 'Expert systems: Where are we? And where do we go from here?' The AI Magazine, 3, 3-22.

Davis, R. (1984) Amplifying expertise with expert systems. In: Winston, P.H., Prendergast, K.A. (eds.) The AI business: the commercial uses of artificial intelligence. Cambridge, Mass : MIT, pp. 17-40.

Davis, R., Buchanan, B.G., Shortliffe, E.H. (1977) 'Production rules as a representation for a knowledge-based consultation program.' Artificial Intelligence, 8, 15-45.

Davies, R., James, B. (1984) 'Towards an expert system for cataloguing: some experiments based on AACR2.' Program, 18, 283-297.

Dennett, D. (1979) Brainstorms: philosophical essays on mind and psychology. Hassocks : Harvester Press.

Denning, P.J. (1986) 'The science of computing: expert systems.' American Scientist, 74, 18-20.

Duda, R.O., Gaschnig, J.G. (1981) 'Knowledge-based expert systems come of age.' Byte, 6, 238-281.

Duda, R.O., Hart, P.E., Reboh, R. (1985) 'Letter to the editor.' Artificial Intelligence, 26, 359-360.

Duda, R.O., Shortliffe, E.H. (1983) 'Expert systems research.' Science, 220, 261-268.

Ernst, G.W., Newell, A. (1969) GPS: A case study in generality and problem solving. New York : Academic Press.

Evanczuk, S., Manuel, T. (1983) 'Practical systems use natural languages and store human expertise.' Electronics, December 1, 139-145.

Feigenbaum, E.A. (1979) Themes and case studies of knowledge engineering. In: Michie, D. (ed.) Expert systems in the micro-electronic age. Edinburgh : Edinburgh University Press, pp. 3-25.

Feigenbaum, E.A., McCorduck, P. (1984) The Fifth Generation: artificial intelligence and Japan's computer challenge to the world. London : Michael Joseph.

Fikes, R., Kehler, T. (1985) 'The role of frame-based representation in reasoning.' Communications of the ACM, 28, 904-920.

Fleck, J. (1982) Development and establishment in artificial intelligence. In: Elias, N., Martins, H., Whitley, R. (eds.) Scientific Establishments and Hierarchies. Sociology of the Sciences Yearbook, Dordrecht : Reidel, 6, 169-217.

Fleck, J. (1984) Artificial intelligence and industrial robots: an automatic end for utopian thought In: Mendelsohn, E., Nowotny, H. (eds.) Nineteen Eighty-Four: Science Between Utopia and Dystopia. Sociology of the Sciences Yearbook, Dordrecht : Reidel, 8, 189-231.

Friedland, P. (ed.) (1985) 'Knowledge-based architectures.' Communications of the ACM, 28.

Genesereth, M.R., Ginsberg, M.L. (1985) 'Logic programming.' Communications of the ACM, 28, 933-941.

Gilbert, G.N., Heath, C. (eds.) (1985) Social action and artificial intelligence. Aldershot : Gower Press.

Goldstein, I., Papert, S. (1977) 'Artificial intelligence, language and the study of knowledge.' Cognitive Science, 1, 84-123.

Gooding, C. (1986) 'Only humans can be expert.' Computer Weekly, May 8, p. 39.

Guilfoyle, Ch. (1986a) 'A table load full of micro shells.' Expert Systems User, 2, 18-20.

Guilfoyle, Ch. (1986b) 'Expert explanation.' Expert Systems User, 2, 25-27.

Hayes-Roth, F. (1985a) 'Knowledge-based expert systems - the state of the art in the US.' Knowledge Engineering Review, 1, 18-27.

Hayes-Roth, F. (1985b) 'Rule-based systems.' Communications of the ACM, 28, 921-932.

Hayes-Roth, F., Waterman, D.A., Lenat, D.B. (1983) An overview of expert systems. In: Hayes-Roth, F., Waterman, D., Lenat, D.B. (eds.) Building expert systems. Reading, MA : Addison-Wesley, pp. 3-29.

Hjerppe, R. (1983) What artificial intelligence can, could, and can't do for libraries and information services. Proceedings of the 7th IOLIM, Dec 6-8, 1983, London : Learned Information, pp. 7-25.

Hjerppe, R., Olander, B., Marklund, K. (1985) Project ESSCAPE - Expert Systems for Simple Choice of Access Points for Entries: applications of artificial intelligence in cataloguing. (Paper presented at IFLA 51st General Conference in Chicago, 18-24 August, 1985).

Johnston, R. (1986) 'ART and KEE: are they worth the price?' Expert Systems User, 1, 10-11.

Kornhauser, A. (1983) Teaching chemistry today: what do we aim for? In: Grunewald, H. (ed.) Chemistry for the future. Oxford : Pergamon Press, pp. 383-392.

Kornhauser, A. (1985) Razvoj v smeri inzeniringa znanja (Towards knowledge engineering). (1. posvetovanje Sekcije za specialne knjiznice Zveze bibliotekarskih drustev Slovenije - Paper presented at the 1st Conference of the Section for Special Libraries of the Library Association of Slovenia, Ljubljana, 12 November, 1985).

Kowalski, R.A. (1977) Predicate logic as a programming language. New York : North-Holland.

Kraft, A. (1984) XCON: an expert configuration system at Digital Equipment Corporation. In: Winston, P.H., Prendergast, K.A. (eds.) The AI business: the commercial uses of artificial intelligence. Cambridge, Mass : MIT, pp. 41-49.

Kuhn, T.S. (1962) The structure of scientific revolutions. Chicago : University of Chicago Press.

Leith, P. (1986) 'Why the experts must be accountable.' Computer Guardian, May 8, p. 15.

Lenat, D.B. (1977) On automated scientific theory formation: a case study using the AM program. In: Hayes, J.E., Michie, D., Mickulich, L.I. (eds.) Machine Intelligence 9. New York : Halsted, pp. 251-286.

Lenat, D.B. (1982) 'The nature of heuristics.' Artificial Intelligence, 19, 189-221.

Lenat, D.B. (1984) 'Computer software for intelligent systems.' Scientific American, 251, 152-160.

Lindsay et al. (1980) Applications of artificial intelligence for organic chemistry: The DENDRAL project. New York : McGraw-Hill.

McDermott, J. (1982) 'R1: a rule-based configurer of computer systems.' Artificial Intelligence, 19, 39-88.

Manchester, P. (1986) 'Clubs take over cult of artificial intelligence.' Computer Weekly, May 8, p. 38.

Marcus, R. S. (1983) 'An experimental comparison of the effectiveness of computers and humans as search intermediaries.' Journal of the American Society for Information Science, 34, 381-404.

Michard, A. (1982) 'Graphical presentation of boolean expressions in a database query language.' Behaviour and Information Technology, 1, 279-288.

Michie, D. (1984) 'Automating the synthesis of expert knowledge.' Aslib Proceedings, 36, 337-343.

Michie, D., Johnston, R. (1985) The creative computer: machine intelligence and human knowledge. Harmondsworth : Penguin.

Minsky, M.L. (1979) Computer science and the representation of knowledge. In: Dertouzos, M.L., Moses, J. (eds.) The computer age: a twenty-year view. Cambridge, Mass : MIT, pp. 392-421.

Mozetic, I., Bratko, I., Lavrac, N. (1983) An experiment in automatic synthesis of expert knowledge through qualitative modelling. Proc. Logic Prog. Workshop, Albufeira, Portugal, June 26 - July 1, 1983.

Mylopoulos, J., Levesque, H.J. (1983) An overview of knowledge representation. In: Brodie, M., Mylopoulos, J., Schmidt, J. (eds.) On conceptual modelling. New York : Springer-Verlag, pp. 3-17.

Nau, D.S. (1983) 'Expert computer systems.' Computer, 16, 63-85.

Newell, A. (1983) Intellectual issues in the history of artificial intelligence. In: Machlup, F., Mansfield, U. (eds.) The study of information. New York : Wiley, pp. 187-227.

Newell, A., Simon, H.A. (1972) Human problem solving. Englewood Cliffs, N.J. : Prentice-Hall.

Nilsson, N.J. (1971) Problem-solving methods in artificial intelligence. New York : McGraw-Hill.

Nilsson, N.J. (1982) Principles of artificial intelligence. Berlin : Springer.

Partridge, D. (1986) Social implications of artificial intelligence. In: Yazdani, M. (ed.) Artificial intelligence: principles and applications. New York : Chapman and Hall, pp. 315-336.

Pogson, B. (1986) 'Emulated.' Computer Guardian, May 22, p. 17.

Polanyi, M. (1967) The tacit dimension. New York : Anchor.

Pollitt, A.S. (1981) An expert system as an online search intermediary. In: Proceedings of the 5th International Online Information Meeting. Oxford : Learned Information, pp. 25-32.

Pollitt, A.S. (1984) 'A "front-end" system: an expert system as an online search intermediary.' Aslib Proceedings, 36, 229-234.

Pollitt, A.S. (1986) An expert systems approach to document retrieval: a summary of the CANSEARCH Research Project. Huddersfield Polytechnic, (Technical Report Series, no 86/6).

Quinlan, J.R. (1979) Discovering rules by induction from large collections of examples. In: Michie, D. (ed.) Expert systems in the micro-electronic age. Edinburgh : Edinburgh University Press, pp. 168-201.

Ritchie, G.D., Hanna, F.K. (1982) AM: a case study in AI methodology. Edinburgh : University of Edinburgh (Department of Artificial Intelligence, Technical Report No. TR18).

Robinson, J. (1965) 'A machine oriented logic based on the resolution principle.' Journal of the ACM, 12, 23-41.

Schank, R.C. (1975) Conceptual information processing. New York : North Holland.

Searle, J. (1980) 'Minds, brains and programs.' Behavioural and Brain Sciences, 2, 417-457.

Searle, J. (1983) Intentionality: an essay in the philosophy of mind. Cambridge : Cambridge University Press.

Sheil, B. (1983) 'Power tools for programmers.' Datamation, February, 131-144.

Shortliffe, E.H. (1976) Computer based medical consultations: MYCIN. New York : American Elsevier.

Small, H. (1984) The lives of a scientific paper. In: Warren, K. (ed.) Selectivity in information systems. New York : Praeger, pp. 83-97.

Small, H. (1986) 'The synthesis of specialty narratives from co-citation clusters.' Journal of the American Society for Information Science, 37, 97-110.

Smith, L.C. (1976) 'Artificial intelligence in information retrieval systems.' Information Processing and Management, 12, 189-222.

Smith, L.C. (1985) Knowledge-based systems, artificial intelligence and human factors. (Paper prepared for the Seminar on Information Technology as a Tool for Information Use, Copenhagen, 8-10 May, 1985).

Smith, L.C. (1986) Machine intelligence vs. machine-aided intelligence as a basis for interface dasign. (Paper presented at 2nd Conference on Computer Interfaces and Intermediaries for Information Retrieval, 29 May, Boston, 1986).

Sowizral, H.A. (1985) 'Expert systems.' Annual Review of Information Science and Technology, 20, 179-199.

Stamper, R. (1985) Knowledge as action: a logic of social norms and individual affordances. In: Gilbert, G.N., Heath, C. (eds.) Social action and artificial intelligence. Aldershot : Gower Press, pp. 172-191.

Stefik, M. (1981) 'Planning with constraints (MOLGEN Part 1).' Artificial Intelligence, 16, 111-140.

Stefik, M. (1981) 'Planning and meta-planning (MOLGEN Part 2).' Artificial Intelligence, 16, 141-170.

Town, W.G. et al. (1985) Advanced computational support for biotechnology research: project report. Prepared by Hampden Data Services, U.K., on behalf of Commission of the European Communities ITTTF/CUBE, May 1985.

Turing, A. (1950) 'Computing machinery and intelligence.' Mind, 59, 433-460.

van Melle, W. (1980) A domain-independent system that aids in constructing knowledge-based consultation programs. Stanford Heuristic Programming Project Memo HPP-80-22.

Vickery, B.C. (1968) 'Bibliographic description, arrangement and retrieval.' Journal of Documentation, 24, 1-15.

Waldrop, M.M. (1984) 'Artificial intelligence (I): into the world.' Science, 223, 802-805.

Walker, D.E. (1981) 'The organization and use of information: contributions of information science, computational linguistics and artificial intelligence.' Journal of the American Society for Information Science, 32, 347-363.

Waltz, D.L. (1982) 'Artificial intelligence.' Scientific American, 247, 101-122.

Ward, S.A., Reed, L.J. (eds.) (1983) Knowledge structure and use: implications for synthesis and interpretation. Philadelphia : Temple University Press.

Webster, R., Miner, L. (1982) 'Expert systems: programming problem-solving.' Technology, 2, 62-73.

Weizenbaum, J. (1985) Computer power and human reason: from judgment to calculation. Harmondsworth : Penguin.

Welbank, M. (1983) A review of knowledge acquisition techniques for expert systems. Martlesham Heath : British Telecom Research Laboratories.

Winograd, T. (1972) Understanding natural language. New York : Academic Press.

Winston, P.H., Prendergast, K.A. (eds.) (1984) The AI business: the commercial uses of artificial intelligence. Cambridge, Mass : MIT.

Woolgar, S. (1985) 'Why not a sociology of machines? The case of sociology and artificial intelligence.' Sociology, 19, 557-572.

Yaghmai, N.S., Maxin, J.A. (1984) 'Expert systems: a tutorial.' Journal of the American Society for Information Science, 35, 297-305.

Zadeh, L.A. (1979) A theory of approximate reasoning. In: Hayes, J.E., Michie, D., Mickulich, L.I. (eds.) Machine Intelligence 9. New York : Wiley.